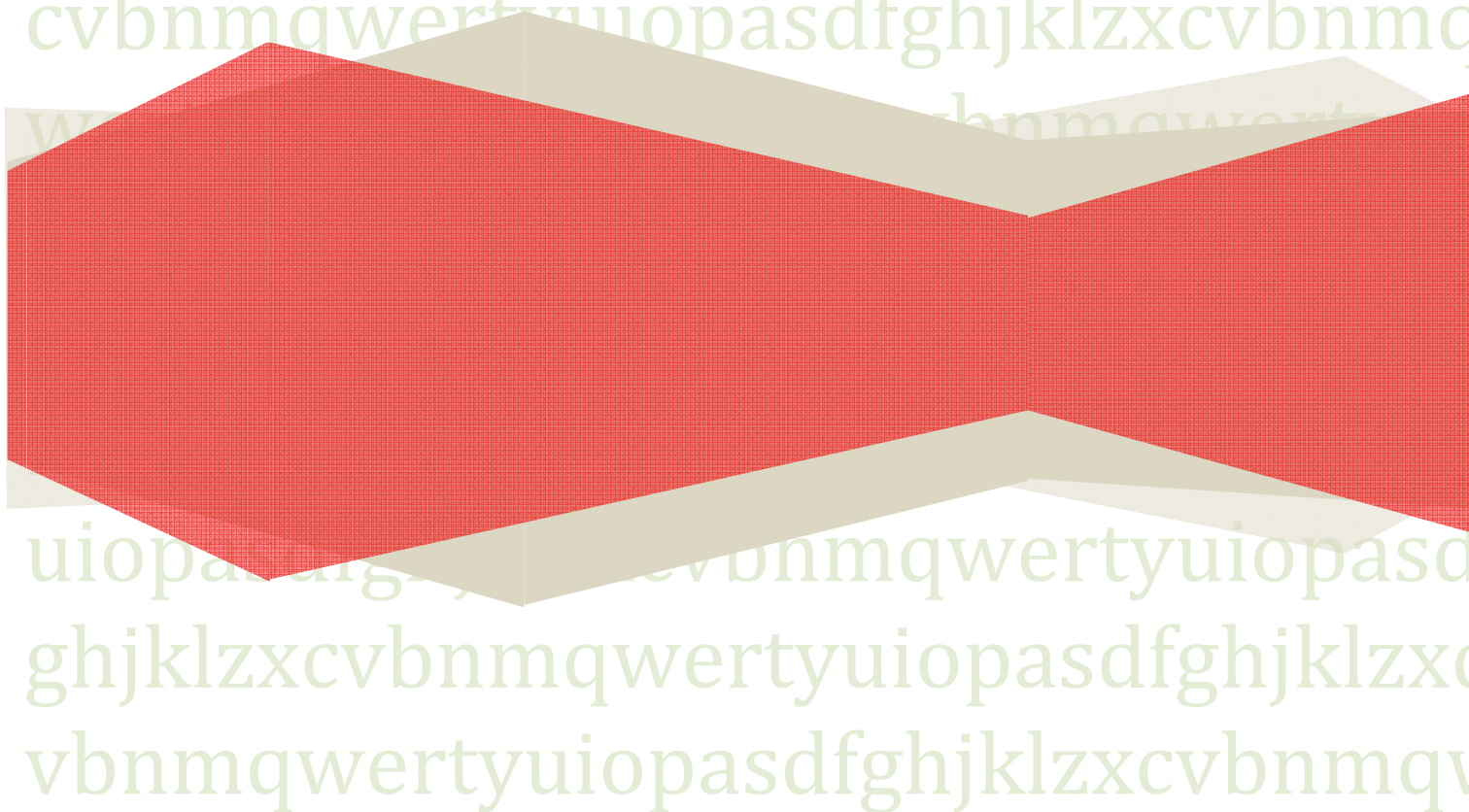Jolanta Dadonienė
Kęstutis Žagminas
Aušra Beržanskytė

Introduction to research methodology

Jolanta Dadonienė

Kęstutis Žagminas

Aušra Beržanskytė

Introduction to research methodology

Content:

Chapter 1. Basics concepts in epidemiology

Causation and causal inference

Measures of disease frequencies

Measures of effect

Measures of effect in the exposed population

Measures of effect in the total population

Causation and causal inference

A good research sense is elusive without training in epidemiology concepts and methods.

The word epidemiology comes from the Greek words *epi*, meaning on or upon, *demos*, meaning people, and *logos*, meaning the study of. Many definitions have been proposed, but the following definition captures the underlying principles and public health spirit of epidemiology: the study of the occurrence and distribution of health related events in specified populations, including the study of the determinants influencing such events, and the application of this knowledge to control health problems. The study includes surveillance, observation, hypothesis testing, analytic research, and experiments. *Distribution* refers to analysis by time, place, and classes or subgroups of persons affected in a population or in a society. *Determinants* are all the physical, biological, social, cultural, economic and behavioral factors that influence health. *Health-related states and events* include diseases, causes of death, behaviors, reactions to preventive programs, and provision and use of health services. *Specified populations* are those with common identifiable characteristics. Application to control makes explicit the aim of epidemiology—to promote, protect, and restore health. The primary "knowledge object" of epidemiology as a scientific discipline are the causes of health-related events in populations (Porta et al. 2008). The principal objective of epidemiology is to study the causation of health-related events or conditions in humans. Epidemiology focuses on the question of general causation: is the agent capable of causing a disease? A cause is an act or event or state of nature which initiates or permits, alone or in conjunction with other causes, a sequence of events resulting in an effect (Rothman

Key words:
ratio
proportion
rate
incidence
risk
incidence proportion
attack rate
incidence rate
prevalence
risk difference
relative risk
odds ratio
attributable risk
attributable fraction
preventable fraction
population attributable risk
population attributable fraction

1976). Rothman has elaborated a component cause model that attempts to accommodate the multiplicity of factors that contribute to the occurrence of an outcome. In his model, a sufficient cause is represented by a complete circle (a "causal pie"), the segments of which represent component causes. When all of the component causes are present, then the sufficient cause is complete and the outcome occurs.

The concept of cause must be distinguished from the notion of association. An association is not equivalent to causation. Assessing whether an association is causal requires an understanding of the strengths and weaknesses of the study's design as well as judgment about how the study findings fit with other scientific knowledge. In assessing causation epidemiologists first look for alternative explanations for the association, such as bias or confounding factors.

A series of logical, empirical, and theoretical checks that causal relations may or may not be satisfying was developed by Austin Bradford Hill in 1965(HILL 1965) and elaborated by others (Susser 1991). They are:

"Consistency: The association is consistent when results are replicated in studies in different geographical and time settings, or in studies conducted by different methods. Evidence of a similar relationship between a factor and disease in various groups is considered to be more suggestive of a causal role for the factor than evidence of variation in the association.

Strength: This is defined by the size of the risk as measured by appropriate statistical estimates. Large relative risks or odds ratio are more likely to be causal, although weak relationships may also be causal.

Specificity: It means that the exposure is associated with a particular disease and not with disease in general. This criterion has its origins in infectious disease epidemiology. This criterion may be quite inappropriate in non-infectious disease epidemiology. For example, cigarette smoke is involved in the pathogenesis of a wide range of outcomes.

Dose-response relationship: An increasing level of exposure (in amount and/or time) increases the risk. For example, the number of deaths due to lung cancer is proportional to the number of cigarettes smoked per day.

Temporal relationship: Exposure always precedes the outcome. This is probably the only indisputable criterion for causality.

Biological plausibility: This refers to the agreement of the examined association with existing biological knowledge.

Coherence: This means that the association is not conflicting with scientific knowledge on the disease in terms of biologic, physical, and social mechanisms

Experiment: This refers to the possibility of eliciting the outcome by experimentally introducing exposure or as the possibility of preventing the outcome by removing the exposure."

There is no algorithm that can be used to asses whether a causal inference is appropriately based on these viewpoints. One or more factors may be absent even when a true causal relationship exists. Similarly, the existence of same factors does not ensure that a causal relationship exists.

Drawing causal inferences after finding an association and considering these factors requires judgment based on biology (Green MD, Freedman DM, Gordis L 2000). It is important to keep in mind that most judgments of cause in epidemiology are tentative and should remain open to change with new evidence. Checklists of causal criteria should not replace critical thinking.

Measures of disease frequency

We have defined epidemiology as the study of the occurrence and distribution of health-related events in specified populations. Other sciences, such as clinical medicine, are also directed toward the study of health and disease, but in epidemiology the focus is on population distributions. The objective of most epidemiologic research is to obtain a valid and precise estimate of the effect of a potential cause on the occurrence of disease. To achieve this objective, an epidemiologist must be able to measure the frequency of disease occurrence, either in absolute or in relative terms (Rothman, Greenland & Lash 2008).

Common frequency measures are *ratios, proportions,* and *rates.* All three frequency measures have the same basic form:

$$\frac{\text{Numerator}}{\text{Denominator}} \times 10^n$$

A **ratio** is the relative magnitude of two quantities or a comparison of any two values. The numerator and denominator need not be related.

$$\frac{\text{Number of events, persons etc. in one group}}{\text{Number of events, persons etc. in another group}}$$

In certain ratios, the numerator and denominator are different categories of the same variable, such as males and females. In other ratios, the numerator and denominator are completely different variables, such as height and weight. The numerators and denominators of a ratio can be related (the number of males and the number of females in the population) or unrelated (the number of physicians in the city and size of the population living in that city).

A **proportion** is the comparison of a part to the whole. It is a type of ratio in which the numerator is included in the denominator. You might use a proportion to describe what fraction of patients are seropositive for HIV, or what percentage of the population is smokers.

$$\frac{\text{Number of persons with a particular characteristic}}{\text{Total number of persons of which the numerator is a subset}} \times 10^n$$

In a proportion, the numerator must be included in the denominator. For a proportion, 10^n is usually 100 (or n=2) and is often expressed as a percentage.

A ***rate*** is a measure of the frequency with which an event occurs in a defined population over a specified period of time. Rate means how fast something is happening or going. In epidemiology a rate describes how quickly disease occurs in a population.

In epidemiology the main measures of disease frequency are: *risk, incidence* rate and *prevalence.*

Risk is a probability that an event will occur, e.g. that an individual will become ill or die within a stated period of time or by a certain age. Risk is an incidence proportion. Incidence proportion is a proportion because the persons in the numerator, those who develop disease, are all included in the denominator (the entire population).

$$\text{Risk} = \frac{\text{Number of new cases of disease during specified time period}}{\text{Number of persons followed for the specified time period}}$$

Incidence proportion is the proportion of an initially disease-free population that develops disease, becomes injured, or dies during a specified (usually limited) period of time. Synonyms include risk, cumulative incidence, attack rate. The measure of risk requires that all of persons (initially disease-free) are observed for entire time period during which the risk is being measured. Incidence proportion is a measure of risk. Therefore the time duration must be specified for it to be meaningful. The only way to interpret a risk is to know the length of the time period over which the risk applies. Without identifying the time period risk values are not meaningful. In the outbreak setting, the term ***attack rate*** is often used as a synonym for risk. It is the risk of getting the disease during a specified period, such as the duration of an outbreak. The time reference for an attack rate is usually not stated but implied by the biology of the disease being described.

Incidence rate (person-time rate) is similar to incidence proportion in that the numerator is the same. It is a number of new cases that occur in a population. However, the denominator is different. The denominator is the sum of the time each person was observed, totaled for all persons. This denominator represents the total time experienced for the subjects observed. Thus, the incidence rate is the ratio of the number of cases to the total time the population is at risk of disease.

$$\text{Incidence rate} = \frac{\text{Number of new cases of disease during specified period}}{\text{Time each person was observed, totaled for all persons at risk of disease}}$$

An incidence rate describes how quickly disease occurs in a population. One important concept is that incidence rate, like speed, is an instantaneous concept. An incidence rate is the momentary rate at which cases are occurring within a group of people.

Incidence rate is based on person-time, so it has some advantages over an incidence proportion. Because person-time is calculated for each subject, it can accommodate persons coming

into and leaving the study. The denominator accounts for study participants who are lost to follow-up or who die during the study period. Epidemiologists commonly calculate annual incidence rates based on a numerator of cases observed and a denominator based on the mid-year population. This type of incident rate turns out to be comparable to a person-time rate.

Prevalence proportion (sometimes referred to as prevalence rate) does not measure disease onset. Prevalence proportion or simply prevalence is the proportion of persons in a population who have a particular disease or attribute at a specified point in time or over a specified period of time.

$$\text{Prevalence } = \frac{\text{All new and pre - existing cases during a given time period}}{\text{Population during the same time period}} \times 10^{n}$$

Incidence proportion and incidence rate are measures that assess the frequency of disease onset. Prevalence differs from incidence in that prevalence includes all cases, both new and preexisting, in the population at the specified time, whereas incidence is limited to new cases only.

Several factors affect prevalence. The greater incidence of disease, the more people will have it. Prevalence is also related to the duration of illness. If disease lasts a short time its prevalence is lower than if it lasts a long time. Diseases with short duration may have a low prevalence even if the incidence rate is high. Since prevalence can be influenced by incidence rate and disease duration it is not as useful as incidence for studying the causes of diseases. Prevalence is useful for measuring the disease burden on a population.

Measures of effect

The principal objective of epidemiology is to study the causes of disease. However, the first question is whether an association exists between exposure and disease. Observed association does not necessarily mean that there is a causal relationship. Causal relationship is only one of possible explanation. Person may be exposed to an agent and then develop disease without there being any causal relationship between exposure and disease. "To measure a causal effect we have to contrast the experience of exposure people with what have happened in the absence of exposure" (Rothman 2002). Measure of effect is a theoretical concept, since such two measurements, in the same group of persons under study, are not feasible during the same time period. In order to approach this theoretical situation as closely as possible, we will use as unexposed group a population similar to the exposed group but for the exposure. In these two populations (or in 2 subsets of the same population, exposed and unexposed), we will then measure and compare disease occurrence. So, we can directly observe an association instead causal effect. If the two groups are not equal, then the measure of association will not equal the measure of effect. In such a circumstance, we say that the measure of association is confounded.

A measure of association is a statistical concept that quantifies the effect or potential effect of an exposure on a disease. For example, is there a greater frequency of lung cancer in the population

of smokers than nonsmokers? If so, it might be inferred that smoking is a risk factor for lung disease. In epidemiology it is common to use the term exposure to denote any explanatory variable i.e. we may speak of smoking as an exposure that causes lung cancer. If we observe that two variables (disease and exposure) are related in some way, we refer to them as being associated.

It is common to present information about these variables in a table form.

Table 1.1. Example of a 2x2 table

	Disease +	Disease -	Total
E+	a	b	a+b
E-	c	d	c+d
Total	a+c	b+d	a+b+c+d

The most common measures of association are: ***difference measures*** and ***ratio measures.*** Risk difference would be the difference in incidence proportion or risk between the exposed and unexposed groups. With an incidence rate instead of risk we can calculate the incidence rate difference. In cross sectional studies, the difference measure is called the prevalence difference, and is estimated as the difference between two prevalence estimates. Difference measures of association cannot be estimated in case-control studies because in such studies neither risks, rate, prevalence can be appropriately estimated. For example, ***risk difference*** (RD) is the absolute difference in risk associated with exposure:

$$RD = R_1 - R_0$$

R_1 represents the risk in the exposed group = $\dfrac{a}{a+b}$

R_0 represents the risk in the non-exposed group = $\dfrac{c}{c+d}$

If exposure and getting the disease is not associated the difference between incidence rates would be close to zero. A rate difference different from zero indicates an association, but usually the subtraction is performed so that the rate difference is positive.

The second method of measuring association is more popular among epidemiologists. Rather than subtracting the incidences, they are divided, giving what is known as the ***relative risk*** (RR). Term *relative risk* is applied to incidence rate ratio and risk (incidence proportion) ratio and sometimes to odds ratio as well (Rothman, Greenland & Lash 2008). The relative risk is defined as follows:

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{R_1}{R_0} \quad \text{(Risk ratio)}$$

$$RR = \frac{I_1}{I_0} \quad \text{(Rate ratio)}$$

I_1 – incidence rate in exposed cohort
I_0 – incidence rate in non-exposed cohort

The odds ratio is the odds of disease, given exposure. The ***odds ratio*** (OR) is interpreted in the same way as risk ratio. The odds ratio is calculated as:

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

When the number of cases of disease is relatively low if compared to the number of non-cases (i.e. the disease is rare), then the odds ratio approximates risk ratio. If the risk of disease is relatively low in both exposed and non-exposed individuals, then *a* will be small relatively to *b* and *c* will be small relative to *d*. As a result:

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} \cong \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc} = OR$$

A risk difference close to 0.0—that is, incidence rates or risks close to each other for the exposed and non-exposed populations—indicates a lack of association between getting a disease and being exposed. Relative risk, close to 1.0 would indicate a lack of association. Why the relative risk is more popular than the rate difference? Rate difference can mask important differences in differences. For example, the rate difference from risks of 95% and 90% and the rate difference from risks of 10% and 5% are both 5%. Their similarity masks distinctions that are very important to epidemiologists. The relative risk is also easily interpreted with reference to possible risk factors. A relative risk close to 1.0 would indicate that the incidences for the exposed and non-exposed populations are about the same. A relative risk (or odds ratio) greater than 1.0 would indicate that exposure may well be a risk factor, and a relative risk (or odds ratio) less than 1.0 would indicate exposure to a factor that seems to decrease the risk of (preventive factor) (Rothman KJ 2002).

Measures of effect in the exposed population

Epidemiology is not just about identifying risk factors for disease but also about evaluating interventions to reduce or eliminate the effect of these risk factors. It is therefore important to predict the impact of removing a particular risk factor on the incidence of disease in the population. Therefore, we need a way of measuring the proportion of the disease that can be attributed to the exposure.

If we take the risk difference (RD) between risk in exposed (R_1) and risk in unexposed (R_0) people we obtain the absolute quantity of the outcome measure that is associated with exposure. This is known as the ***attributable risk*** among the exposed (ARe).

$$ARe = RD = R_1 - R_0$$

Attributable risk is defined as the increase or decrease in the risk of disease in the exposed group that is attributable to exposure. It assumes that the causal effect is entirely due to the risk factor. If we divide risk difference by the risk in exposed group we obtain ***attributable fraction*** (AFe).

$$AFe = \frac{RD}{R_1} = \frac{R_1 - R_0}{R_1} = 1 - \frac{1}{RR} = \frac{RR - 1}{RR}$$

Attributable fraction is the proportion of disease in the exposed group that is due to exposure. It assumes that risk difference reflects a causal effect that is not distorted by any bias.

If exposure prevents disease (e.g. vaccination), the attributable risk is often called the ***preventable fraction*** among the exposed (PFe).

$$PFe = \frac{R_0 - R_1}{R_0} = 1 - RR$$

For case-control studies, if odds ratio approximates relative risk, then attributable fraction can be approximated:

$$AFe = \frac{OR - 1}{OR}$$

This approximation is appropriate if controls are representative of the general population and the prevalence of exposure is low.

Measures of effect in the total population

Population attributable risk (PAR) is the increase or decrease in risk of disease in the population that is attributable to exposure. Using the notation defined above, population attributable risk (PAR) is calculated as:

$$PAR = R_{Total} - R_0$$

What proportion on the disease experience in the whole population is attributable to a particular exposure? The measure that answers this question is known as the population attributable fraction. Population attributable fraction is the proportion of disease in the population that is due to the exposure (Rockhill, Newman & Weinberg 1998). Using the notation defined above, the *population attributable fraction* (PAF) is calculated as:

$$PAF = \frac{R_{Total} - R_0}{R_{Total}}$$

If disease is rare over time interval, ratio of incidence rates (I_0/I_{Total}) approximates the ratio of risks (incidence proportions) and thus formula can be written as:

$$PAF = \frac{I_{Total} - I_0}{I_{Total}}$$

When there is no confounding of exposure-disease association:

$$PAF = \frac{p_e(RR - 1)}{p_e(RR - 1) + 1}$$

p_e – proportion of source population exposed to the factor of interest.

If confounding exists PAF can be is calculated as:

$$PAF = \frac{P_e(RR - 1)}{RR}$$

P_e – proportion of cases exposed to risk factor.

We suggest following the data based on findings of the Rotterdam study that was presented as an example by Kleinbaum DG et al. in 2003 (Table 1.2) (Kleinbaum,D.G., Sullivan K.M., Barker N.D. 2003). The rationale for the study evolved from the presumption that hypothyroidism, a disease state in which the production of thyroid hormone is decreased, is known to increase the risk of cardiovascular disease. The Rotterdam study investigators therefore examined the potential effect of subclinical hypothyroidism on the incidence of myocardial infarction in the study population.

Table 1.2. 2x2 table for population attributable fraction calculation in Rotterdam study

	MI		Total
	Yes	No	
Subclinical hypothyroidism Yes	3	100	103
Subclinical hypothyroidism No	10	844	854
Total	13	854	957

In this study of nearly 1 000 women aged 55 and over were examined for thyroid hormone levels initially and followed for a certain time until the outcome.

To calculate the attributable fraction (*AFe*) we need to know the risk (= incidence proportion) in women with subclinical hypothyroidism, and in women without hypothyroidism, namely R_1 and R_0.

R_1 represents the risk in the group with hypothyroidism = $\dfrac{a}{a+b}$ = 2.9%

R_0 represents the risk in the group without hypothyroidism = $\dfrac{c}{c+d}$ = 1.2%

The attributable fraction $AFe = \dfrac{RD}{R_1} = \dfrac{R_1 - R_0}{R_1} = 1 - \dfrac{1}{RR} = \dfrac{RR-1}{RR}$ = 59%

Thus in the women that are affected by subclinical hypothyroidism, almost 60% of the myocardial infarction can be attributed to the presence of subclinical hypothyroidism (attributable fraction).

Population attributable fraction (PAF) can also be calculated from this example. Taking in account the formula for population attributable fraction

$$PAF = \frac{R_{Total} - R_0}{R_{Total}}$$

the total risk R_{Total} should be calculated beforehand. In the presented example it is 13/957 = 1.4%.

The population attributed fraction in this example is: *PAF* = (1.4 − 1.2) / 1.4 = 14%. This indicates that of all myocardial infarctions that occur in elderly women, 14% are due to the presence of subclinical hypothyroidism. In other words, if subclinical hypothyroidism could be prevented, there would be 14% less myocardial infarctions in this population.

Can we calculate the population attributed fraction using the alternative formula given above in this lecture? Yes, we can and the result you get is completely the same.

$$PAF = \frac{p_e(RR-1)}{p_e(RR-1)+1}$$

In this equation p_e – proportion of source population exposed to the factor of interest, or simpler a prevalence of hypothyroidism in this population 103/957 = 0.108 or 10.8%.

RR = (3/103) / (10/854) = 2.5

$$PAF = \frac{0.108(2.5-1)}{0.108(2.5-1)+1} = 14\%$$

Interpretation of population attributed fraction assumes that the rate in the exposed group would decline to that of nonexposed group if the exposure was eliminated. Population attributable fraction is a function of the strength of the association and the prevalence of the exposure in the population. The prevalence of exposure can decline or increase. Therefore, the population attributable fraction is specific to a particular time and place.

Chapter 2. Error in Epidemiological Studies

Understanding of a valid measure of interest
Random error
Systematic error
Selection bias
Information bias
Preventing bias
Confounding
Interaction

Understanding of a valid measure of interest

The principal objective of epidemiology is to study the causation of health-related events or conditions in humans. Epidemiologic studies can be viewed as exercises in measurement, with the objective to measure the frequency of disease in population or effect of exposure as accurately as possible. Therefore, the primary objective of most epidemiologic research is to obtain a valid estimate of an effect measure of interest.

Suppose a study is conducted to measure the ratio of the incidence rate (RR) of myocardial infarction (MI) and physical activity. We can imagine that there is a correct value (parameter) of RR. A given study will produce an estimate of this correct value ($\overline{RR}$). If the study estimates a value $\overline{RR}$ close to RR, we would consider the study to be accurate, so it has little error. If study estimate $\overline{RR}$ differs significantly from correct RR value we would consider the study to be inaccurate. This estimate is affected to both random and systematic error. Therefore, there are two different sources of inaccuracy that can occur when estimating an effect: *systematic error* (validity problem) and *random error* (precision problem). Also we should to take into account that epidemiologic estimate is the end product of the study design, study conduct and data analysis. All these steps of the estimation process are associated with error. We often think of precision and validity as separate ideas but, in fact, they are related to each other.

Key words:
systematic error
random error
estimate
precision
bias
validity
target population
study population
external validity
selection bias
information bias
confounding
interaction

Estimate = parameter + random error + systematic error

The random and systematic errors inherent in an estimate bring the value of the estimate up or down from the parameter by varying amounts. For example, $\overline{RR}$ = 0.60 = [0.3(parameter) +0.2(random error) + 0.1 (systematic error)].

Random error

"Random error is the portion of variation in a measurement that has no apparent connection to any other measurement or variable, generally regarded as due to chance" (Porta et al. 2008). In an epidemiologic study, random variation has many sources. There are three main levels of variability: individual, population and sample. Variability within an individual can occur because of biological changes in the individual over time (e.g., heart rate, blood pressure, total serum cholesterol level, etc.), changes related to factors such as age, diet, environmental factors.

Another source includes measurement error. Measurements errors can occur in all three levels of variability (individual, population and sample). There are three main sources of random error in making measurements: observer variability, instrument variability and subject variability.

There is also variability in population which can be thought of as the cumulative variability of individuals. Because populations are made up of individuals with different genetic constitutions, who are subject to different environmental influences, populations often exhibit more variation than individuals.

Another component is sampling variation. When epidemiologists perform analytic studies, they cannot usually study the entire population. Instead they study subsets or samples of the population. This introduces another source of variability—sampling variability. Using a single sample of subjects to represent the population is analogous to using a single measurement to characterize an individual. Repeated samples from the population will give different estimates of the true population values.

Random and systematic errors are components of sampling error. Both random and systematic errors can also contribute to measurement error.

Precision is defined as a relative lack of random error. In statistics, one measure of precision is the inverse of the variance of a measurement or estimate. A measure of imprecision is the standard error of measurement.

A common way to increase precision or reduce random error is to increase size of the study. Other ways to reduce random error and increase the precision of measurements are: standardizing measurement methods, training and certifying observers, refining instruments, automating the instruments, repetition (Hulley 2007).

Systematic error

Systematic error is also called bias. ***Bias*** can be defined as systematic deviation of results or inferences from truth. It is "an error in the conception and design of a study—or in the collection, analysis, interpretation, reporting, publication, or review of data and leading to results or conclusions that are systematically (as opposed to randomly) different from truth" (Porta et al. 2008).

"Ways in which deviation from the truth can occur include:

1. Systematic variation of measurements from the true values (measurement error).
2. Variation of statistical summary measures (means, rates, measures of association, etc.) from their true values as a result of systematic variation of measurements, other flaws in study conduct and data collection, flaws in study design, or analysis.
3. Deviation of inferences from truth as a result of conceptual or methodological flaws in study conception or design, data collection, or the analysis or interpretation of results.
4. A tendency of procedures (in study design, data collection, analysis, interpretation, review, or publication) to yield results or conclusions that depart from the truth.
5. Prejudice leading to the conscious or unconscious selection of research hypotheses or procedures that depart from the truth in a particular direction or to one-sidedness in the interpretation of results" (Porta et al. 2008).

The opposite of bias is validity. ***Validity*** is relative absence of bias or systematic error. Validity and precision are both components of ***accuracy***.

The validity of a study is usually separated into two components: *internal validity* and *external validity*. Terms "external validity", "internal validity" depend on the population to which inferences are made.

External population ---- Target (source) population -----Study population ---- Sample

Statistical inference

Internal validity

External validity
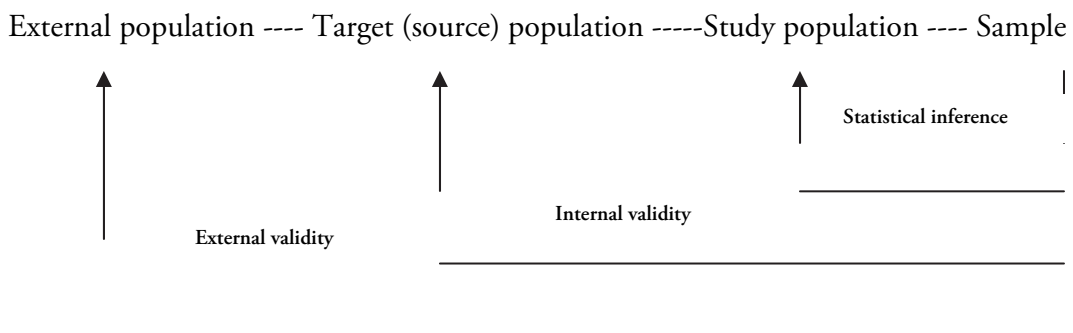
The population about which we wish to draw conclusion is called ***target (source) population***. This population is the collection of individuals of restricted interest from which one has sampled and about which one wishes to make inferences with respect to study objective. In many cases this population is defined according to geographical, institutional or occupational criteria (specific city, community, occupation). The specific population from which data are collected is called the ***study population***. This is eligible population. ***External population*** is the collection of individuals to

which the study has not been restricted, for example, a different city, community, or occupation, but to which one still wishes to generalize the study conclusion.

We can make statistical inferences from the sample to the study population, but we would like to make inferences from the sample to the target population. Due to methodological features of the study design, the study population may not be representative of the target population due to systematic error.

With regard to the populations defined above the term ***internal validity*** concerns the validity of inferences about the target (source) population using information from study population. Internal validity depends on methods used to select the study subjects, collect information and conduct analysis. ***External validity*** (generalizability) is the extent to which the results of a study are applicable to other populations. Internal validity parallels the statistical concept of generalizing from sample to source population, while generalizability involves more informal inference beyond a source population to external populations.

Of the two types of validity, internal validity is more important than external validity, since it does not make sense to generalize findings that are not internally valid. As a result, internal validity needs to be evaluated carefully before one even considers the external validity of a study.

The validity of a study is dependent on the degree of systematic error. Epidemiologists frequently classify systematic error into three broad categories: ***selection bias*** (bias in the way the study subjects are selected), ***information bias*** (bias in the way the study variables are measured), and ***confounding***.

Selection bias

Selection bias refers to a systematic error in a study resulting from the manner in which the subjects are selected or retained in the study. This error can occur when the characteristics of the subjects selected for a study differ systematically from those in the target population or when the study and comparison groups are selected from different populations.

Selection bias may be due to: sampling bias, ascertainment bias (case ascertainment (surveillance) bias, referral / admission bias, and diagnostic bias), participation bias (self-selection, volunteerism), healthy worker effect, non-response / refusal bias, survival bias, loss to follow-up.

Selection bias can occur in a cross sectional study when a convenience sample (as opposed to a probability sample) is employed.

In a case-control study, controls should be drawn from the same population as the cases, so they are representative of the population which produced the cases. Controls are used to provide an estimate of the exposure rate in the population. Therefore, selection bias may occur when those individuals selected as controls are unrepresentative of the population that produced the cases. The potential for selection bias in case-control studies is a particular problem when cases and controls are recruited exclusively from hospital or clinics. However, hospital patients tend to have

different characteristics to the wider population, for example they may have higher levels of alcohol consumption or cigarette smoking. Their admission to hospital may even be related to their exposure status, so measurements of the exposure among controls may be different from that in the reference population. This may result in a biased estimate of the association between exposure and disease.

Selection biases in case-control studies include among others: surveillance bias, diagnostic bias, non-response bias, survival bias. Surveillance bias happens when there is more intense surveillance orscreening for the outcome among the exposed than among the unexposed. Diagnostic bias can occur when the diagnostic approach is related to knowledge of the subject's prior exposure to a putative cause (e.g. taking a certain drug, being exposed in an outbreak etc.). Non-response bias is a systematic error due to the differences in response rates of participants in a study. Survival bias occurs when survivors of a highly lethal disease are more likely to enter a study than other cases.

Selection bias can also occur in cohort and experimental studies. Selection bias is less of a problem in cohort studies compared with case-control studies, because exposed and unexposed individuals are enrolled before they develop the outcome of interest. Selection biases in cohort studies include: healthy worker effect, diagnostic bias and loss to follow-up.

The healthy worker effect is a potential form of selection bias specific to occupational cohort studies. For example, an occupational cohort study might seek to compare disease rates amongst individuals from a particular occupational group with individuals in an external standard population. There is a risk of bias here because individuals who are employed generally have to be healthy in order to work. In contrast, the general population will also include those who are unfit to work. Therefore, mortality or morbidity rates in the occupation group cohort may be lower initially than in the population as a whole.

Diagnostic bias can also occur in cohort studies if the diagnosis depends on the knowledge of the exposure status. Selection bias may be introduced when the completeness of follow-up or case ascertainment differs between exposure categories. For example, it may be easier to follow up exposed individuals who all work in the same factory, than unexposed controls selected from the community. This can be minimized by ensuring that a high level of follow-up is maintained among all study groups.

Randomized trials are theoretically less likely to be affected by selection bias, because individuals are randomly allocated to the groups being compared, and steps should be taken to minimize the ability of investigators or participants to influence this allocation process. However, refusals to participate in a study, or subsequent withdrawals, may affect the results if the reasons are related to exposure status.

In brief, selection bias can occur in any type of epidemiologic study, although it is more common in case-control studies because case and controls are often selected on the basis of different criteria, which in turn may be related to the frequency of exposure. Selection bias is problematic because it can result in an overor underestimation of the true magnitude of the relationship between an exposure and an outcome. The potential for selection bias is not always easily recognized. To

avoid selection bias investigators need to be very careful that the study and comparison groups are similar except for the variables being investigated, and that subject losses are kept to a minimum.

Information bias

Information bias results from systematic differences in the way data on exposure or outcome are obtained from the various study groups (Hennekens et al. 1987). The term "misclassification" is frequently used to describe this bias. Misclassification results in an incorrect estimation of the association between exposure and outcome, the size and direction of this depending on the type of misclassification of exposure or outcome. The misclassification can be differential or non-differential.

Differential misclassification occurs when the extent of misclassification is different between the study and comparison groups. For example, in a case-control study if more cases are mistakenly classified as being exposed than controls, then the misclassification is differential. Similarly, in a cohort study if the exposed group is more likely to be mistakenly classified as having developed the outcome than the unexposed group, then again the misclassification is differential.

Nondifferential misclassification occurs when the degree of misclassification between study and comparison groups is uniform, that is, when there is an equal frequency of incorrect classifications on exposure status among those with and without the exposed. (Rothman 2002). It is important to distinguish between differential and nondifferential misclassification since they produce different effects on the measures of association. Like selection bias, differential misclassification leads to over- or underestimation of the true magnitude of the measure of association. If the cases in a case-control study are more likely to be misclassified as being exposed compared to the controls, then the study will likely overestimate the magnitude of the odds ratio (i.e. produce positive bias). If, on the other hand, the controls are more likely to be misclassified as being exposed than the cases, the study will likely underestimate the magnitude of the odds ratio (i.e. produce negative bias). Unlike differential misclassification, nondifferential misclassification results in a dilution of the measure of association. That is, the measure of association is biased toward the null value, the value that represents no association between the exposure and outcome (e.g. a relative risk of one or a risk difference of zero). Therefore, nondifferential misclassification can result in the apparent absence of an association when in fact an association exists. While differential misclassification can lead to either underestimation or overestimation of an association, nondifferential misclassification invariably biases the association toward the null value.

Misclassifications might be introduced by the observer (interviewer bias, biased follow-up), by the study participants (recall bias, prevarication), or by measurement tools such as questionnaires or instruments such as weighing scales or blood pressure cuffs.

Observer bias occurs when there are systematic differences in the way information is collected for the groups being studied. This may be a result of the investigator's prior knowledge of the hypothesis under investigation or knowledge of an individual's exposure or disease status. Such

information may influence the way information is collected, measured or interpretation by the investigator for each of the study groups.

Loss to follow up is a particular problem associated with cohort studies. Bias may be introduced if the individuals lost to follow-up differ with respect to the exposure and outcome from those persons who remain in the study.

Recall bias may occur when the information provided on exposure differs between the cases and controls. In a case-control study data on exposure is collected retrospectively. The quality of the data, therefore, is determined to a large extent on the patient's ability to accurately recall past exposures. For example an individual with the outcome under investigation (case) may report their exposure experience differently than an individual without the outcome (control) under investigation. That is, cases may tend to have a better recall on past exposures than controls. Recall bias may result in either an underestimation or overestimation of the association between exposure and outcome.

Preventing bias

Selection and information biases are best controlled by prevention during the design, data collection, and execution phases of a study. This means that potential sources of bias must first be recognized as potential threats to the internal validity of a study. Once they are recognized, measures must be taken to see that their potential is minimized to the extent possible. This generally will be easier to accomplish in experimental than observational studies, but the goal must always be to minimize bias in a study. Various procedures have been developed to minimize different types of bias. These include using randomly selected samples where possible, standardizing measurement instruments and protocols, using objective means of verifying exposures and outcomes (e.g. laboratory tests), blinding investigators as to the status of study subjects on either exposure and/or outcome as appropriate, and aggressively following up subjects who withdraw from a study so as to determine their exposure and outcome status.

Confounding

Most epidemiologic studies are designed to estimate the effect of some exposure factor or factors on the risk of disease. Even when an association exists epidemiologists must determine whether the exposure causes the disease or apparent association between exposure and disease actually is due to another variable. Alternatively, the apparent lack of an association could result from failure to control for the effect of some other factor. While the results of an epidemiological study may reflect the true effect of an exposure on the development of the outcome under investigation, it should always be considered that the findings may in fact be due to an alternative explanation. Such alternative explanations may be due to the effects of chance (random error), bias or confounding

which may produce spurious results, leading us to conclude the existence of a valid statistical association when one does not exist or alternatively the absence of an association when one is truly present (Hennekens et al. 1987). Observational studies are particularly susceptible to the effects of chance, bias and confounding and all three need to be considered at both the design and analysis stage of an epidemiological study, so their potential effects can be minimized.

Classical approaches to confounding focus on individual variables. A confounding factor is a variable that is associated with the exposure and, independently of exposure, a risk factor for the outcome, but not caused by the exposure. It creates a bias in the estimated association between an exposure and an outcome as a result of its associations with exposure and outcome. Confounding can be viewed as the mixing of the effect of the exposure under study on the outcome of interest with that of a third variable (or a group of variables), i.e., the confounder(s).

If a factor is associated with exposure but not outcome, or is associated with outcome but not exposure, it will not be a confounder.

Following conditions are necessary for extraneous variable to be confounder: 1) must be associated with the exposure among the source population for cases; 2) it must be associated with disease among unexposed individuals; 3) it should not be an intermediate variable in the causal pathway under study (Rothman, Greenland & Lash 2008).
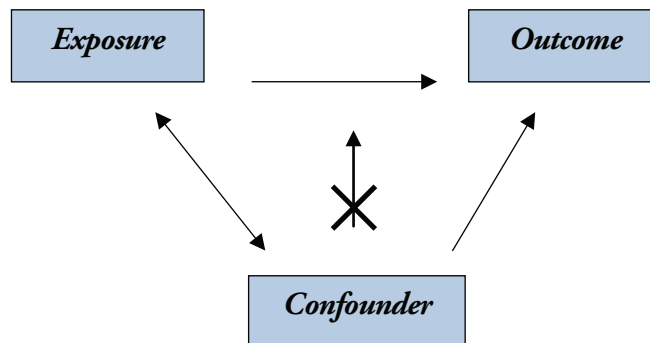
Like bias, confounding represents systematic error and threatens the internal validity of an epidemiologic study since it can lead to false conclusions regarding the true relationship between an exposure and outcome. Confounding can either overestimate or underestimate the true magnitude of the measure of association between an exposure and outcome. When the effect of a confounder overestimates the magnitude of a measure of association, it is said to be a positive confounder. Conversely, if the confounding factor leads to underestimation of the magnitude of the measure of association, it is said to be a negative confounder. Depending on the nature of its relationship with the exposure and outcome, a confounder can even distort associations to such an extent that a positive association appears negative or no association appears as an association.

Because a confounding factor is one that is independently associated with both the exposure and the outcome, one way of assessing confounding is to evaluate each of those associations. However,

because confounding is related only to the magnitude (i.e., the strength) of those associations, and not on their statistical significance, it is inappropriate to use statistical tests or *p*-values to assess confounding ((Lang JM, Rothman KJ, Cann CI. 1998). Perhaps the most widely used statistical test of confounding is a comparison of baseline characteristics among exposure groups. When the sample size is large, small imbalances in baseline characteristics across exposure groups might be statistically significant, even though there would be minimal confounding. Conversely, when the sample size is small, large baseline differences may not be statistically significant. Reliance on the statistical test would lead the investigator to conclude, falsely, that those baseline characteristics could not be confounders. Tests of the associations between potential confounders and the outcome, as is done when using automated selection procedures such as stepwise, forward, or backward, are also inappropriate for the selection of confounders. Those procedures are driven by *p*-values, which reflect both the strength of the association between a potential confounder and the outcome, as well as the size of the sample, whereas confounding is related only to the strength of that association. Automated selection procedures should be used only when the researcher's aim is to create prediction models or risk scores for the outcome. A more appropriate way to assess whether an individual variable is a confounder is to determine whether the unadjusted measure of the association between the exposure and the outcome differs from the measure, adjusted for that variable, by a meaningful amount. Unfortunately, there are no clear guidelines for what constitutes a meaningful difference, though in practice, with ratio measures of effect (such as OR or other measures of the relative risk) many epidemiologists use a 10% to 20% change in estimate criterion ((Kurth T 2007).

There are several ways to deal with confounding, some simple, others more complicated. They all assume that two conditions are satisfied: 1) all confounders have been identified or at least suspected; 2) identified or suspected confounders can be adequately conceptualized and accurately measured (Greenland 1980).

Confounding can be prevented in the design or controlled in the analysis of a study. In randomized trials with sufficient sample size, randomization prevents confounding on average, by creating groups that are essentially identical (exchangeable) at baseline. In any given trial, however, there may be imbalances across exposure groups in confounding factors due to chance alone. In observational studies, restriction or matching on exposure factors can be used to prevent confounding. By restricting the study population to one level of a confounder (e.g., one gender), one removes the association between that confounder and the exposure. Similarly, by matching on several confounders, the association between these confounders and exposure is removed. If the factors that are being restricted or matched are measured with error, there may be residual confounding despite restriction or matching. There are several ways of controlling for confounding in the analysis. *Standardization* controls confounding by the application of a standard distribution of confounding variables to all exposure groups. *Stratification* controls confounding by analysis of the exposure/outcome association in homogeneous subgroups, or strata of confounders. Like matching, stratification is unwieldy with more than a few variables due to the number of

strata required. *Multivariable regression* models are the most commonly used method to control for confounding because such models allow simultaneous control for multiple confounders. In multiple regression models, the exposure and confounders are included as independent variables in the model so that such models control for confounding while estimating the effect of exposure on the outcome (Kurth T 2007).

Interaction

Confounding and interaction are different phenomena. Interaction term has been used to describe different biological and statistical concepts. Interaction can be defined as the interdependent operation of two or more causes to produce, prevent, or control an effect. "*Biological interaction* means the interdependent operation of two or more biological causes to produce, prevent, or control an effect"(Porta et al. 2008).

From a biological point of view, component causes within the same sufficient cause may be thought of as interacting biologically. In other words, the exposures act synergistically to produce disease, since in the absence of one factor, disease will not occur by that mechanism. From epidemiological point of view, interaction is frequently characterized as effect modification: factor A and factor B alone have certain relationship with a disease, but together the factors have a different effect. This effect may be greater or smaller than expected based on the magnitude of the individual effects(Adami, Hunter & Trichopoulos 2008).

Chapter 3. Types of epidemiological research studies

Classification of research studies
Descriptive studies
Analytic studies
Hybrid designs

Classification of research studies

The classification of research studies though complicated from the first glance may be simply divided into two broad groups, namely quantitative and qualitative studies although this chapter and the book in general deals with quantitative research. The quantitative studies can be divided into descriptive and analytic research studies and the discrimination between them is the key point for understanding the differences between the quantitative studies in general. There is no unified classification for different biomedical sciences and it slightly differs depending on the objectives it covers. The following classification is adapted from the one proposed by the Centre of Evidence Based Medicine located at University of Oxford the broad aim of which is to develop, teach and promote evidence-based health care and provide support and resources to doctors and health care professionals to help maintain the highest standards of medicine and public health[1] (Figure 3.1).

Descriptive studies

A ***descriptive study*** does not make an attempt to quantify or estimate a relationship but seeks to provide us a picture of what is happening in a population or group, be it patients or whatever group of interest, e.g., the prevalence of disease or just a frequency of some symptom within the patient group, or just an experience of a group. The key elements of the descriptive studies are the place, time and group of people. A descriptive study may have different

Key words:
quantitative studies
qualitative studies
descriptive studies
cross sectional studies
survey/prevalence studies
correlational studies/ecological studies
case reports/case series
experimental study
observational analytic studies
clinical trials
case control study
cohort study
exposure
outcome
surveillance study
case-cohort study
nested case-control study

---

1   http://www.cebm.net (CEBM>EBM Tools>Critical Appraisal>Study Designs)

subtypes depending on whether it is performed on a small group of patients or large populations and different geographical areas. So far, descriptive studies include:

- **cross sectional (or survey/prevalence) studies,**
- **correlational studies (ecological studies),**
- **case reports or case series.**

If the study is related to small group of inhabitants or patients the most common name is cross sectional study. In cross sectional study the disease status, often the prevalence of disease, are assessed simultaneously among individuals in a well-defined population. The cross sectional studies may sometimes also include analytic work when comparing the distribution of factors between the groups and hence the size of the problem, and applying more complex statistical methods than rather descriptive statistical analysis. Note, it never gives the clue for causation, describing the association at most. If the descriptive epidemiological study is performed in a large population and examines the differences in disease rates among populations in relation to age, gender, race, and differences in temporal or environmental conditions it is called *survey or prevalence study*. There can be a specific time window, such as a given calendar year during which a community-wide survey is conducted (Levin 2005, Mann 2003).

In *correlational* studies, measures that represent characteristics of entire populations or big groups are used to describe disease occurrence in time in relation to some factors of interest, like product consumption, smoking habits, economic growth, etc. The groups may be census tracts, states, or countries. The correlation coefficient, denoted by r, is the important descriptive measure of association in correlational studies. This coefficient quantifies the extent to which there is a linear relationship between exposure and disease. That is, the basic data are typically percentages or other summary statistics rather than measurements of characteristics on individuals, or measures of effect. A chief strength of correlational studies, which contributes to their frequent use, is that they can be done quickly and inexpensively, often using already available information. Governmental health agencies routinely collect demographic and product consumption data that can be correlated with disease incidence, mortality, or utilization of health resources. Similarly, the availability of data from surveillance programs or national and international disease registries can permit comparisons of disease rates in different geographic regions. This is why they are sometimes called ecological studies. The chief limitation of correlational studies is the inability to link exposure with disease in particular individuals. In particular, a correlational study has the data on the number of exposed persons and the number of cases but does not have the number of exposed cases. For example, there is a strong positive correlation between coronary heart disease mortality rates and cigarette sales per capita, but there is not enough data to prove that those who buy cigarettes, tend to smoke and die more frequently. A second major limitation of correlational studies is the lack of ability to control for the effects of potential confounding factors. Finally, correlational data represent average exposure levels rather than actual individual values. Showing overall positive or negative linear

association, this might actually be masking more complicated relationship between exposure and disease. In general, these studies can only identify patterns or trends in disease occurrence over time or in different geographical locations but cannot ascertain the causal agent or degree of exposure. These studies are often very useful for generating hypotheses for further research.

While correlational studies consider whole population, *case reports* describe the experience of a single patient or a group of patients with a similar diagnosis. These types of studies, in which an unusual feature of disease or patient history is reported, may lead to formulation of a new hypothesis and represent an important link between epidemiology and clinical medicine. They account for one third of publications in medical journals. One fundamental limitation of the case report is that it is based on experience of only one or few persons. The presence of any risk factors, however suggestive may simply be coincidental. Another, that case reports have only minor relevance to public health.

Analytic studies

Much more complicated design stands within analytic design of research studies consisting of two different types of studies:
- **experimental**
- **observational analytic**

Experimental arm of this branch deals mostly with **clinical trials**. A clinical trial is an experimental study designed to compare the therapeutic or health benefits of two or more treatments. The major objective of a clinical trial is to test the possible effect, that is, the efficacy, of a therapeutic or preventive treatment such as new vaccine, medication, physical therapy, or dietary regimen for either treating or preventing the occurrence of a disease. The long range goal of a *preventive trial* is to prevent disease; the long range goal of a *therapeutic trial* is to cure or control a disease. The objective of most of preventive trials is to assess the effectiveness of a prevention/screening program. Examples of preventive trials include studies of vaccine efficacy, use of aspirin to prevent coronary heart disease, smoking cessation, diet modification, effectiveness of fluoridation, sex education, etc. Therapeutic trials are typically performed by pharmaceutical companies to test new drugs for treatment, although it is not a rule. Therapeutic clinical trials can be designed to compare different methods of rehabilitation or even surgical interventions and are preferred by health specialists.

Experimental research provides data from which firmer conclusions can be drawn if compared with observational analytic part of the current branch. However, it would not always be best to perform a clinical trial. It may be unnecessary when the effect of an intervention is dramatic, e.g. the effect of antibiotics on bacterial wound infection, then the likelihood of an unknown confounding factor may be ignored. Experimentation may be inappropriate when outcomes occur in the distant future and sometimes it takes decades to manifest, e.g. hormone replacement therapy to prevent femoral fracture. Experimentation may not be possible for ethical reasons or political or legal obstacles. It may be also inadequate because of the nature of the procedure e.g. surgery.

On the other hand, often analytic observational studies, in particular **cohort studies** and **case control studies**, are the only practicable method of studying various problems, for example, studies of aetiology, instances where a randomised controlled trial might be unethical, or the condition to be studied is rare. In analytic observational studies, the researcher simply measures the exposure or treatments of the groups. Analytic observational studies investigate and record **exposures** (such as interventions or risk factors) and observe **outcomes** (such as disease) as they occur. Such studies may be purely descriptive or more analytical. Nevertheless, there are the substantial differences between those two study designs and they will be worked out through the content of this book based on information presented in research series of medical journals and websites [2],[3]. The easiest way to distinguish between the studies is to remember how the population is assembled to these studies. The starting point of a cohort study is the recording of healthy subjects with and without exposure to the putative agent (supposed risk factors) or disease characteristic but without the outcome of interest. All individuals with and without exposure are followed the same way and, ideally, the same time period, usually long enough, and their status is observed and recorded during the course of the study. If this kind of study is performed on a substantial part of population with the aim to monitor health problems it is called **surveillance studies**. They focus on continuous monitoring of trends in the occurrence and distribution of disease. A critical component of the definition of surveillance is that surveillance systems include the ongoing collection, analysis, and use of health data. In 1968 the 21st World Health Assembly described surveillance as the systematic collection and use of epidemiological information for the planning, implementation, and assessment of disease control.

As a good example for that may serve WHO MONICA project (Multinational Monitoring of Trends and Determinants of Cardiovascular Disease) focusing on the trends in cardiovascular disease in different populations.

The starting point of a case control study is subjects with the disease or condition under the study (cases). A comparison group consisting of individuals without the disease under study (controls) are assembled. The cases history of exposure or other characteristics, or both, prior to onset of the study are recorded with the same attention to both groups.

For easer remembrance, the cohort study most often, though not necessarily, looks forward (prospective cohort) while the case control study always looks backward (retrospective).

The distinctive feature of analytic observational studies from experimental studies is the role of investigator. While in experimental studies the exposure, namely medications or intervention is prescribed by the investigator in a manner that all other factors should be distributed equally, in observational studies the investigator does not apply or modify exposure (e.g. treatment, intervention, risk factor) and takes it as it is by observing.

---

2 http://www.gfmer.ch
3 http://www.socialresearchmethods.net/kb/index.php

Hybrid design

Although the descriptive, cohort, case control and experimental studies are the key stones of research, different hybrid designs can be met in the medical literature depending on the field of biomedicine, public health, clinical medicine, experimental laboratory research etc. Hybrid designs combine the elements of at least two basic designs. Two popular hybrid designs are the case-cohort study and the nested case-control study. A **case-cohort study** uses a hybrid design that combines elements of a cohort and a case-control study. A case-cohort population is followed over time to identify new or incident cases of a disease. The control group consists of non-cases sampled from the original cohort. As an example, a 1995 study of risk factors for gastric cancer involved a cohort of 9,775 men in Taiwan on whom blood samples were taken and frozen. Subsequent follow-up based on cancer registry data identified 29 cases of gastric cancer. A control group of 220 controls who did not develop gastric cancer was sampled from original cohort. One exposure variable of interest was the presence or absence of Helicobacter pylori infection, which could be assessed by unfreezing and analysing the blood sample from cases and controls. The **nested case-control** study is a variation of the case-cohort study, this type of study can be used if the time at which subjects became cases is known. In this design, controls are matched to cases at the time of cases' diagnosis. In 1993 a nested case-control study designed for cancer risk from serum copper levels, baseline blood specimens and risk factor information that were obtained from 5000 telephone employees. A cancer surveillance system identified 133 cancer cases that developed from this cohort. The time of case-diagnosis was determined and used to choose a sample of 241 controls equal by working time at telephone stations and compared to them with regard to serum copper level and other covariates of interest. So the cases were matched to controls by the working time at telephone stations.

A **qualitative study** explores people's subjective understanding of their lives and experiences. Methods used include direct observation, interviews, the analysis of texts or documents or recorded speech or behavior. A qualitative study can obtain views, opinion and perspectives of individuals included in the study. The examples of this kind of studies may be: the perception of ageing in mass media; the awareness of oral cancer among smokers and drinkers. They may identify themes that could not been identified from a quantitative study. The main disadvantages are that no measures of effect can be reproduced. Second, the qualitative studies are fairly small in size.

The classification itself does not give the clue as to what study design we have to choose in a particular situation and what the result should we seek for. Our first goal when starting the research is to know well enough what we want to explore and, namely, what the question we want to answer is. To raise the proper question and find the right ways to answer it, is the main problem in health research, be it the public health or clinical medicine. It is always worth to go through literature and try to find out whether the same question has already been raised by the others and if so, what additional answers you wish to obtain. Only accurate and exact formulation of research question

leads us to an appropriate research design. The suggestions for questions and answers can be found in Table 3.1.

Table 3.1. How to choose the proper design of a study

If the question of the research study lies in:	he common design is:
Prevalence of disease or Frequency of factor	Cross sectional
Incidence	Cohort
Cause of disease or influence of factors for the outcome (with different reliability)	Cohort, case-control
Prognosis, harm	Cohort
Diagnosis	Diagnostic test studies
Prophylactic and treatment effect	Preventive and therapeutic trial
Effect of community preventive or screening program	Community trials, community intervention trial
Experience of illness	Qualitative

Having raised the study question and defined the study design further questions may arise. Is the study design ethical? The ethical side of the study is decided by the Lithuanian Bioethics Committee together with the investigator. The next question to answer is what resources we have for the planned study. How long does it take to get an answer? How much money will it cost? What personnel should we need for that? Chapter 8 in this book deals with most of these problems. If these problems are manageable, then start the project.

For those who may be keenly interested in the performance and calculations of different types of study we suggest to apply for EPI INFO programme tutorials and to download the data for different type of studies from this website. EPI INFO is a free accessed programme managed by the Center for Disease Control and Prevention, located in Atlanta (USA).

Chapter 4. Cross sectional studies

Cross sectional design
How to perform a cross sectional study
Advantages and disadvantages of cross sectional study

Cross sectional design

Cross sectional/observational studies are primarily used to determine prevalence of the outcome of interest for a given population, commonly for the purposes of public health resources planning. Prevalence equals the number of cases in a population at a given point in time. Not to forget, the cohort studies are used to determine the incidence of the outcome of interest in a given population. In cross sectional studies all the measurements on each individual are made at one point in time (Figure 4.1). The point prevalence is the most common measure of disease, which is defined as the probability that an individual in a population is a case at that point in time.

Key words:
cross sectional studies
point prevalence
prevalence difference
sample size
thumb rules
random sample
stratified sample
systematic sample
cluster sample
questionnaire

$$\text{Prevalence} \ = \frac{\text{All new and preexistin g cases during a given time period}}{\text{Population during the same time period}} \times 10^n$$

For detailed explanation of this measure of frequency refer to Chapter 1.

This way the cross sectional studies provide a "snapshot" of the outcome and the characteristics associated with it, at a specific point in time. The cross sectional studies are non-directional (directionality) since the subjects are classified on exposure and health outcome status on the same time. And they are retrospective (timing) since the health outcome occurred prior to the study.

How to perform a cross sectional study (survey)

The performance of a cross sectional study can be generally structured into necessary steps to be performed to get valid results. Though it may seem rather simple, it is quite complicated in reality when it comes to detailed planning and protocol writing. For details please refer to Chapter 7.

1. Establish the purpose and specific tasks of the study. This is the concrete tasks to be fulfilled and this way to answer less rigorous purpose.
2. Decide on the sample size: the number of people to interview and examine
3. Create or obtain the questionnaire: the validation of the questionnaire falls under strict rules
4. Data collection method: most often interview method
5. Conduct the interview
6. Analyze the data: the result you get is often limited to the frequencies

The *purpose* of the study is descriptive, often in the form of a survey or a simple questionnaire. The goals of the project determine to whom the survey will be addressed and what questions will be asked. If the goals are unclear, the results will probably be unclear. Usually there is no hypothesis as such, but the aim is to describe a population or a subgroup within the population with respect to an outcome and a set of risk factors. They are limited, however, by the fact that they are carried out at one time point and give no indication of the sequence of events, i.e. – whether exposure occurred before, after or during the onset of the disease outcome. At one point in time the subjects are assessed to determine if they have the outcome and risk factor, some of them will have both, some of them only risk or outcome and some of them will not have any. Only associations can be investigated to the most without predicting the causality. Nevertheless, cross sectional studies indicate associations that may exist and are therefore in generating the hypothesis for further research (Levin 2006a).

The result received from this kind of studies is anticipated to be generalized to the population as a whole or the study results can properly characterize the situation in the population. It is so when the **sample size** used in the study has been calculated before the study and the sample is selected from the population by random or alternative valid technique. This way it is likely that the population can be well represented and the results of the study can be generalized to the population. Otherwise a small sample will reflect the group from which it is drawn but not a population as a whole. There is number of computer based methods or simple formulas to count the sample size depending on the level of confidence, expected prevalence of outcome and precision. We suggest using both Open Epi[4] or *Statcalc* from Epi Info menu for sample size estimation for prevalence studies and case control studies. Open Epi Info *utilities* drop down menu and choose *Statcalc*. Run *sample size & power* for *population survey*. You will be asked to enter three figures: the size of population from which the sample will be selected, expected frequency of the factor under study and the worst acceptable result. If your task would be counting the prevalence of the disease, e.g. rheumatoid arthritis in Vilnius, you are supposed to enter the necessary data: the population of Vilnius inhabitants – 600 000; expected prevalence of the disease – 0.5% and the worst acceptable result (precision that satisfies your needs) – e.g. 0.3%. You can choose a different worst acceptable result as it all depends on how exact result you expect to receive. According to the confidence level

---

4 http://www.openepi.com/OE2.3/Menu/OpenEpiMenu.htm

you can choose the sample size for this prevalence study. The confidence level of 95% is sufficient to indicate that the sample size of 4 740 is required for this study.

The alternative way on deciding about the sample size is the so called **rule of thumb**. In general, thumb rules lack scientific basis and many scientists dislike them. When no baseline information for computation of a sample size is available and the pilot study is not manageable, the following rules can be used. For a descriptive study that seeks to find normal levels in a healthy population at least 200 subjects are required. For analytic observational case control study a smaller sample size is needed and can be up to 100 in each group. While in cohort study the number to be followed up should be big enough to ensure that at least 30 persons are finally available with the outcome of interest in each group with exposure and without. A large-sized clinical trial should include nearly 300 subjects in each group, a mid-sized trial nearly 100 in each group, a small-sized trial at least 30 in each group. The last can be used for post graduate thesis where the time and resources are limited (Indrayan 2008).

Having counted the needed sample size the next step is to assemble the responders into the sample in the right way. First of all one should be able to distinguish among eight types of samples that fall into two main categories: *probability* samples and *nonprobability* samples. Four subtypes of probability samples are called a simple **random sample** and **stratified random sample, systematic sample and cluster sample**. **Random samples** and *randomization* are two different issues. One should not mix up these concepts. The presence of **random** in both names suggests that they both involve the use of a probability device. With **random samples**, chance determines who will be in the sample. With *randomization*, chance determines the assignment of treatments or other interventions and it is the case for experimental studies and not for the other types of the studies.. A **random sample** is drawn from a population by using a probability device. Suppose you require a sample of one hundred persons to be surveyed out of one thousand you may access. It would be a mistake to take the first one hundred consecutive persons. The right way to choose the set of one hundred persons is to choose them randomly. It is not always you possess the list of persons to get the random sample and it is never available when you target the sample out of population. You may then apply to the Population Register Database which is managed by Residents Register service in Lithuania and the required random sample of residents, as a whole or in strata's, can be ordered and purchased not confront the legal acts.

A **stratified sample** is a random sample for a subgroup. It is sometimes useful to establish strata's to ensure that your sample accurately reflects relevant sub-groups in your target population. For example, men and women may differ in terms of the outcome you are looking for. If you want your study to accurately reflect the same outcome in the population, you will want to ensure that the percentage of men and women in your sample reflect the percentage of the general population. If not, the results will hardly be applicable to the population. Even if you have managed with gender differences properly, another variable, age, may be distort the results from the truth. We should

introduce correction for such a potentially misleading effect. One popular method of making such a correction is the direct rate adjustment. If the confounding factor is age, this method is generally called age-adjustment, and the corrected rates are called age-adjusted rates. The goal of age adjustment is to modify the crude rates so that the results you get (e.g. prevalence of rheumatoid arthritis in Vilnius) cannot be explained by differences of age between your sample and standard population of Vilnius. The confounding factor, age, is removed by re-computing the rates substituting a common age distribution for the separate age distributions. The common age distribution is determined by identifying a standard population. The actual calculations of the age adjusted rates are not shown here and will be worked out during the classes using Excel statistic package facilities.

The third type of probability sample is called a **systematic sample.** This type of a sample is created when the researcher goes through an ordered list of members of the population and selects for example, every fifth entry on the list (e.g. telephone book) to be in the example. So long as the starting position on the list is determined randomly, each entry on the full list has an equal chance of appearing in the sample.

The last of the four kinds of probability samples involves the so called **cluster samples**. The cluster can be households, schools, universities or any other grouping of people that make one community. Next, a sample of these clusters is randomly selected, for example three schools out of all Vilnius schools. Finally, data are collected from randomly selected persons that are in each of the cluster, namely school.

The *nonprobability* samples are those where the researcher applies the inclusion criteria for the individuals to be involved into the study. This is always the case in clinical trials when certain groups of people with a certain disease are examined. The non probability samples include *purposive samples, convenience samples, quota samples and snowball samples.*

Once you have decided on your sample size and how to organize it, you must then decide on your **questionnaire** or other instrument. Usually we use the questionnaires that were already created for the same purposes we want them to use.

Whatever type of questionnaire you choose, keep it short and simple unless you are stacked to the original one. If you present a 20 page questionnaire most potential responders will give up in horror before even starting. Having designed and validated your questionnaire (read the tips for questionnaires in the end of this section) decide on your *data collection method*. Each method has advantages and disadvantages. *Personal interview* is the most appreciated because answers you get are the most accurate and faithful and sometimes longer interviews are tolerated. Personal interviews usually cost more per interview than other methods. This is particularly true of in-home interviews, where the traveler time is a major financial factor.

Survey by telephone is the most popular interviewing method all over the world because telephone coverage is almost universal. People can usually be contacted faster over the telephone than with other methods and it is relatively cheap. Nevertheless many people are reluctant to answer phone interviews especially sensitive questions. On the other hand it is usual that the phone call is

picked up by the children or women in a household eliminating or substantially diminishing the participation of men in a survey.

Mail surveys are among least expensive but this is the kind of survey where the addresses and names are needed. It always takes longer time to conduct the survey. It also needs another envelope for response to be placed inside the first one.

Email surveys are both very economical and very fast. There is practically no cost and you can get enormous amount of answers during the short time. There are substantial disadvantages also. First, you must possess a list of email addresses; second you cannot use email answers to generalize findings to the whole population, finally many people dislike unsolicited email and delete it as a rule. This way you may want to send email surveys only to people who expect to get email from you.

Bias can occur in any research and reflects the potential that the sample studied is not representative of the population it was drawn or the population at large. The cross sectional design is also very prone to bias when compared to other designs and from the beginning of the study. The level of non-response is the first concern in mail surveys. In populations of lower education and literacy levels, the response rates to mail surveys are often too small to be useful. This, in effect, eliminates substantial part of population. Even in well-educated populations, response rates vary from as low as 3% up to 90%.

A numbers of techniques is employed to minimize the nonresponse including telephone call first with the letter following the call, second and third mailing of surveys, letters emphasizing the importance of response, etc. Another concern is that of biased response, where the person is more likely to respond when they have a particular characteristic or set of characteristics. Bias will occur when the characteristic in question is in some way related to the probability of having the outcome.

Our experience with surveys comes from two studies, conducted in 2006 and 2007, and both were exploring the prevalence of chronic rheumatic disease, systemic lupus erythematosus and rheumatoid arthritis, respectively.

Ten thousand inhabitants in the first study were selected following the calculation of required sample size of 8 000 and expecting the response of about 80%. It is a pity, there were only 40,2 % of respondents out of 10 000 who returned the questionnaires back despite a secondary mailing and a return envelope inside the letter as a techniques employed to raise the number of respondents. Therefore we emphasized that the low response rate may be important when considering the final result, in particular the lupus prevalence.

The prevalence of systemic lupus erythematosus in Lithuania: the lowest rate in Northern Europe. The aim of this study was to explore the prevalence of systemic lupus erythematosus (SLE) in Lithuania (Vilnius). Two different studies were designed for SLE cases identification: registry-based SLE study and population-based SLE study. For the registry-based study patients were enrolled during the period of 1999–2004 and from two sources, including out-patient clinics of Vilnius and tertiary rheumatology center with interview during the year 2004. Only

Vilnius residents who fulfilled the ACR 1982 revised criteria for the classification of SLE were counted in this study. Seventy-six living adult patients with SLE were interviewed and accounted for the prevalence of 16.2/100000 (0.016%) using the Vilnius adult population in January 2004 (a population of 470451). The population study of randomly selected 10,000 Vilnius inhabitants with beforehand validation of the survey was performed in the same year. The population-based study revealed two cases for 4017 respondents, but the low response rate may be important. Extrapolating the results to population of 10000 inhabitants, the point prevalence of SLE in the entire sample was at least 0.02%. Therefore, the prevalence of SLE in Lithuania is the lowest if compared to Northern European countries **Lupus 2006;15(8):544-6**

The second study was more successful whereas the telephone survey was used for rheumatoid arthritis and spondyloarthropathies prevalence study. Although there were a lot of telephone numbers that never responded, disconnected or refused to participate and some business numbers, the response rate of 62.5% and 67.7% in obtaining full interview was reached out of 3370 and 3172 in Vilnius and Kaunas, in respect. We considered it a cost and time saving survey to be conducted in the community.

Prevalence survey of rheumatoid arthritis and spondyloarthropathy in Lithuania. *To assess the prevalence of rheumatoid arthritis (RA) and spondyloarthropathy (SpA) in two Lithuanian cities, Vilnius and Kaunas. The first step in this study involved the translation and validation of a telephone questionnaire developed by rheumatologists and epidemiologists in France. The second step comprised the prevalence survey. To detect RA and SpA cases in the populations of Vilnius and Kaunas, 6542 subjects selected randomly (every 50th) from the latest telephone book were interviewed by telephone using a validated case detection questionnaire (the screening phase). All subjects with rheumatic symptoms but an uncertain diagnosis were contacted by a rheumatologist (confirmation phase) by telephone. If the diagnosis remained uncertain, the subjects were invited for a rheumatological examination. We attempted to contact 3370 telephone numbers in Vilnius and 3172 in Kaunas, and had a response rate of 62.5% and 67.7%, respectively. Over the course of all the study phases (telephone interview, rheumatologist's interview, and clinical examination), 39 RA cases and 27 SpA cases were detected, resulting in a crude prevalence of 0.92% for RA (95% CI 0.65-1.25) and 0.64% (95% CI 0.42-0.92) for SpA. The standardized prevalence rate according to age and sex in the Lithuanian population showed an RA prevalence of 0.55 (95% CI 0.39-0.74) and a SpA prevalence of 0.84 (95% CI 0.53-1.21). The prevalence of RA and SpA in Lithuania was found to be one of the higher rates in Europe. A telephone interview using a validated short questionnaire enabled a cost- and time-saving epidemiological survey to be conducted to detect RA and SpA cases in the community.* **Scand J Rheumatol 2008;37(2):113-9**

There is no benchmark what response rate is acceptable and what we should aim for. It is always worth to remember that if you want a sample of 1 000 people, and you estimate 10% response level, you need then try to contact 10 000 persons.

There are some other disadvantages to cross sectional studies. For example, such a study can identify only existing or prevalent cases at a given time, rather than a new or shortly existing incident cases over a follow-up period. Therefore, it suits best for chronic long standing diseases where the prevalent cases are the best to describe the frequency of the disease. Diseases with short duration, such as the common cold or influenza, especially those that occur during a particular season, may be under-represented by a cross sectional study that looks at the presence of such a disease at a point in time.

The frequency, in particular the prevalence of the outcome, is the main result that is counted and quoted in cross sectional study *analyses*, although sometimes it is erroneously accepted that cross sectional studies do not imply any analysis except frequencies.

Let us follow the study conducted in 1991, in Edinburgh, where a sample of 1592 Scottish men and women between 55 to 74 years were examined for the presence of peripheral vascular disease (Figure 4.2). The study found that the prevalence of intermittent claudication, being a major criteria for peripheral vascular disease, was 72:1592 = 0.045 (4.5%).

Other characteristics, including ischemic heart disease, and smoking were also determined for each subject during the examination. In addition to it, the study found that 5.4% (60 persons) of 1111 ever-smokers had peripheral vascular disease and 2.5% (12 persons) of 481 never-smokers had been diagnosed with this disease. By dividing these frequencies we can see that ever-smokers were 2.16 times more likely to have peripheral vascular disease than never-smokers. This ratio is so called prevalence ratio (PR) that was calculated by dividing two prevalence rates (Table 4.1). Odds ratio can also be calculated from the cross sectional survey to get the better feeling of the risk to contract the disease if smoking. Odds ratio calculation will be covered in detail in the next lecture.

$$PR = \frac{60:1111}{12:481} = 2.16$$

Table 4.1. 2x2 table for prevalence ratio calculation for cross section study

Smoking	Peripheral vascular disease		
	Yes N=72	No N=1520	Total N=1592
Yes	60	1051	1111
No	12	469	481

These results suggested that smoking may contribute to the development of the vascular disease. It is impossible to say , whether vascular disease leads to smoking (as it is extremely painful and smoking may cause a bit of relief from the pain) or smoking leads to vascular disease, which is more reasonable. This illustrates one of the problems with cross sectional studies – they are always non directional.

Advantages and disadvantages of cross sectional study

The most important advantage of cross sectional studies is that they are quick and relatively cheap to perform. The cross section design is a good practice for each researcher starting his carrier to practice on sample size counting, conducting interviews and getting used to statistical packages while performing the simple analysis. As there is no follow up, fewer resources are required to perform the study, though it may take a long period of time to assemble the needed sample of participants. Cross sectional studies are the best way to determine prevalence. The title of the manuscript containing the word prevalence equals to the cross sectional design of the study. They are a useful starting point to extend the study to cohort study or case control study. For cohort study the prevalent cases should be followed up for a substantial period of time, and for case control study – controls can be chosen to correspond the prevalent cases and compared for exposure distribution in both groups.

The most important problem with this type of study is differentiating cause and effect from simple association. These types of studies do not provide explanations for cause effect relation; they do see associations expressed in likelihood ratios to the utmost. Rare conditions cannot be efficiently studied using cross sectional studies because even in large samples there may be no one with the disease. In this situation it is better to study a cross sectional sample of patients who already have the disease (case series).

To sum up the cross sectional study the main advantage of the methos is:
- Cross sectional studies are the best way to study the prevalence of a disease or risk factors
- Quick and cheap

Disadvantages of this type of studies:
- Are mainly descriptive
- Cannot determine incidence
- Cannot be used to assess causality
- Low response rate and bias may occur more often if compared to other design

---

Chapter 5. Case control studies

Case control study design
How to perform a case control study
Advantages and disadvantages of case control studies
Recommendations for using questionnaires in research

Case control study design

Case control study is a basic observational design that starts with cases and non-cases of a disease or other health outcome and proceeds backwards to determine prior exposure history (Figure 5.1). It is retrospective by timing and has got backward directionality. In this study the investigator first selects cases and then chooses controls from persons without the disease/condition. Depending on source population, the case control studies can be population or hospital based studies (Levin 2006c). Note that cases and controls both should come from the same source. In traditional case-control studies controls are selected from people who are still free of the disease at the end of the study period. People with the outcome of interest are compared to the control group who do not have this condition. If the cases and controls are similar by most features except for putative agents, it is called matched case control study. Retrospectively the researcher determines which individuals were exposed to the active agent and which were not. Case control study determines the relative importance of a predictor variable in relation to the presence or absence of the disease.

Key words:
case control study
Creutzfeldt-Jakob Disease
study group
control group
odds ration
source population
matching
confounders
stratifying

How to perform a case control study

1. Decide on the research question to be answered.
2. Decide what variable are most important to be measured and how.
3. Specify the characteristics of the study group and decide how to construct a valid control group.
4. Then compare the exposure of the two groups.

Let's follow the example of **Creutzfeldt-Jakob Disease** (CJD) – a hospital based case control study. CJD is a rare disease characterized by rapidly progressive dementia. In 1990's, a new variant

of CJD in human was discovered in Europe following an epidemic in cattle of mad cow disease, the animal form of CJD. Reacting to this the EU organized a study to investigate whether a diet containing animal products is a risk factor for CJD (van Duijn et al. 1998). The *research question* raised was whether a diet containing animal products is a risk factor for CJD. Since iatrogenic forms of the disease may be linked to animal products the diet information was considered as the most important *variables*. To avoid the information bias it was collected from the patient himself and from his close relative as well. When condition is rare and there is a long latency period the case control study generate a lot of information from relatively few subjects. Consider the practicalities of a cohort study or cross sectional study in assessment of CJD and its aetiology. With less than 300 confirmed cases in Europe a cross sectional study would need about 200 000 subjects to include one symptomatic patient. Taking in account a long latency of the disease, it will take an entire generation to complete the study. Because of that the investigators chose a case control study design and performed the study in two years during 1993–1995. They collected data on 405 cases of **study group** patients with CJD that had occurred in EU. An equal number of control participants were recruited from the hospitals where the patients with CJD had been diagnosed (Figure 5.2).

Sample size determination is also an important point in this kind of study. The following example is referring to the Epi Info tutorial. It depends on exposure level in cases group and controls. Again we suggest using Epi Info statistical package as one of the numerous options. Choose *Statcalc* from *utilities* drop down menu and open *sample size & power for unmatched case-control*. On the screen that appears, accept the default values except for: Expected frequency of exposure in NOT ILL group – insert 50 (50%). Percent exposure among ILL group – insert 60 (60%). With the confidence level of 95% and a power of 80% the sample size required is 407 in each group. If we expect the difference in exposure levels high enough, for example 90% among ILL group and the same 50% in NOT ILL group, than the required number would be only 24 persons in each group and 48 in total. It may happen that we possess just a small sample of, for example, 7 cases with a rare condition then matched case control study would fit better and required sample size would be 1:4, or 7 cases and 28 controls.

In case control studies, data are not available to calculate the incidence rate of the disease being studied, and the actual relative risk cannot be determined. The measure of association between *exposure* and occurrence of disease in case control studies is so-called odds ratio. This measure is described in detail in Lecture 1.

Table 5.1. 2x2 table used for estimation of measure of association in case control studies

Animal products in diet	CJD		Odds of disease in a given exposure
	Yes (cases) N=405	No (controls) N=405	
Yes	a	b	a/b
No	c	d	c/d
Odds of exposure in a disease	a/c	b/d	

The **odds ratio** (OR) or the ratio of odds of disease is thus calculated by

$$OR = \dfrac{\dfrac{a}{b}}{\dfrac{c}{d}} = \dfrac{ad}{bc}$$

It is so called cross product ratio formula because it is the ratio of one product that crosses the table divided by the other product that crosses the table. The rows correspond to whether or not subjects were exposed to the risk factor (consuming animal products) while the columns correspond to being case or control.

Note there is no difference in mathematical result between dividing odds of disease in a given exposure, like it is in example above, or odds of exposure.

$$OR = \dfrac{\dfrac{a}{c}}{\dfrac{b}{d}} = \dfrac{ad}{bc} \ .$$

OR is generally a good estimate of the relative risk and equals it when the disease is rare. Moreover they are interpreted the same way. OR of 1.0 signifies that there is no influence from the exposure variable on the outcome variable. The interpretation of OR is approached in detail in the first lecture. The terms OR and RR are in fact interchangeable when used in context of rare diseases as it is an issue in this CJD study.

Estimate of OR are definitely more popular and quoted more often than RR. It is so because OR estimates association in all kind of studies, namely cross sectional, case control and cohort while RR is an estimate for cohort studies. In other words, if we notice RR in the text of the manuscript it is most probably, the cohort study and if the measure is OR, it may be whatever kind of the study.

Do not forget to calculate the confidence intervals for OR result. Without confidence intervals it is not known whether OR you received are applicable generally or is the result of this study, exclusively. The latter have no scientific meaning in general. The calculation of the confidence intervals is the matter of biomedical statistics and therefore omitted in this book.

Table 5.2. 2x2 table for measure of association in Creutzfeldt – Jakob disease study

Consumption of raw meat	CJD	
	Yes (cases) N=390	No (controls) N=393
Yes	143	106
No	247	187

The case control study for CJD found no evidence for an association between the risk of CJD and the consumption of beef, veal, lamb, cheese, or milk. The few positive findings of the study include increased risk in relation to consumption of brain, and leather products, and fertilizer consisting of hoofs and horns.

$$OR = \frac{a \times d}{b \times c} = \frac{143 \times 187}{247 \times 106} = 1.02$$

From these results you can calculate that among those who had a disease the chance they were eating raw meat was the same as in the control group.

Advantages and disadvantages of case control studies

When a disease or a condition is uncommon, case control studies suit the best and generate a lot of information from relatively small groups. Moreover, a huge number of variables can be considered for putative factors and exposure. This flexibility of the variables comes at the expense of the restricted outcomes studied. The only outcome is the presence or absence of the disease or whatever criteria were chosen to select the cases. The case control study is less prone to bias than cross sectional. But there are still two bias forms that should be thought over before the study begins, in particular sampling bias and information (observer and recall) bias. Both groups, the patients with the disease as well as control group may be biased. Ideally the cases studied should be a random sample of all the patients with the disease. However, to get the list of patients with the disease or condition under the study for random sample to obtain is possible in the countries where disease registries are functioning. In Lithuania the national databases for cancer and tuberculosis provide an opportunity for constructing whatever study and are extremely useful for correlational analysis in time and place, while other registries are not established yet. More frequently health specialists construct the databases for specific condition or disease under interest. Such databases enable vast numbers of people to be entered into a study prospectively or retrospectively. They can be used to construct a cohort, to produce a sample for a cross sectional study, or to identify people with certain conditions or outcomes and produce a sample for a case control study. The real example of the use of the databases is Vilnius rheumatoid arthritis registry which is ongoing since 1998. Nevertheless these databases are a source for selection bias since the patient recruited are those who live in major town of the country and do not fully represent the rural population. Thus you may face the selection bias when the people with some disease/condition are in different status than the same in population.

Selecting valid controls is often more difficult problem. To enable the controls to represent the same population as the cases one of following techniques may be used. First, the ideal control group should be representative of the **source population** and possess the same exposure level as in the source population. As case control studies can be population or hospital based, the controls should be derived from the same source as cases, population or hospital based. Hospital controls are easily accessible, tend to be cooperative, and the interviews are less expensive if compare to population based study. Thus, both cases and controls should come from the same source. The second technique is **matching** – the controls may be matched to cases by several most important

variables that may be the strongest cofounders. How we match cases and controls divides case control studies into two types – the matched and unmatched designs. To qualify as a matched case control study each control must be individually matched, person-to-person, with a case. If cases are independently selected, or are only broadly matched (for example, the same broad mix of ages, same proportions of males and females), then this is an unmatched case control design. With individuals matched person-to person, you have matched paired data, which means that the groups of cases and controls are necessarily the same size. When individuals match each other strictly, the problem of confounding variables is much reduced. However one practical difficulty is that it is sometimes quite hard to find a suitable control to match each of the cases on anything more than age and sex. Third, using two or more control groups may help. If the study demonstrates a significant difference between the patients with the outcome of interest and those without and when the latter have been assembled in a number of different ways, then the conclusion is more robust. For example, in hospital patients can be compared to three different controls, namely, controls from the hospital, and controls from the outpatient clinic and controls form GP patients. It is possible to have up to four controls per case. Another source of bias often met in this kind of studies, is *observer and recall bias*. The researchers are more keen to stress the exposure what they are thinking might be important for the outcome development. To overcome this bias the blinding technique can be employed when the observer doesn't know whom he is interviewing, the case or control. Practically it is impossible and rarely applicable. The same with responders, they tend to put the stress on the things they consider important, neglecting or even forgetting the other ones. To overcome it the data records if available can be looked through (for example, patient records).

Other complicating factors for all observational studies are **confounders**. To avoid confounders we definitely have to ensure that the cases and controls are broadly similar (on age and sex, if nothing else). The reason is that it would be very difficult to identify smoking, say, as a risk factor for lung cancer in the cases, if these were on average twice as old as the control. Used in this context refers to an extraneous variable that satisfies both of conditions: it is a risk factor for the disease being studied, and is associated with the exposure being studied but is not a consequence of exposure. For example, people with asthma are thought to become ill with the lung cancer less frequent than those without asthma. Previously people with asthma were thought to have a "preventive" genetic background until the causal relation of smoking and lung cancer was confirmed in different studies. It came out that patients with asthma tend to smoke less than those without. This way asthma was shown to have relation with outcome (lung cancer) and exposure (smoking) simultaneously. To escape from wrong conclusions and adjust for the effects of confounding factors different techniques can be applied from the beginning of the study. Matching is preferable from the beginning and if not **stratifying** the sample of studying subjects is the second option. Matching for the same frequency of smokers in asthma and not asthma group had to be performed before the study. If not, then analyzing and presenting data for lung cancer frequency those with asthma and

without, separately, is possible. Second, the adjusting for confounders can be performed directly in data analysis by stratified (if there is only one variable supposed to have confounding effect) or multivariate analyses (if several variables may be present).

In summary of case control study the main advantages are:

- Suitable for rare diseases studies with long latencies between exposure and manifestation of disease/condition
- Relatively inexpensive as compared to cohort studies but not to cross sectional
- Can study multiple potential causes of disease
- Can be launched and conducted over relatively short time periods

Main disadvantages:

- Susceptible to selection bias both cases and even more controls
- Susceptible to information bias because information is primarily based on interviews and cannot be validated
- Does not consider more than one disease/condition
- Does not allow estimation of risk though it is sometimes used in term of estimation of influence of exposure

---

Recommendations for using the questionnaires in research

The questionnaires we use may be translated and validated to use for research purposes or may be created *de novo* if there is no such one complying the needs of particular study. The search for already existing questionnaire usually is the better and easier way to obtain the reliable instrument than creating the new one. Moreover it is less costly and less time consuming than generating a new tool. Nevertheless, the cross cultural adaptation of already existing questionnaire also requires careful attention. Usually the questionnaires we use are adapted from English language and should be easily accepted by native speakers not acknowledged with original questionnaire or with problem covered by it. There is a stepwise procedure proposed in the article by Guillemin F. et al. and based on cross cultural adaptation of health related quality of life measures, titled *Guidelines for Cross-cultural Adaptation.*

Guidelines to preserve equivalence in cross-cultural adaptation of health related quality of life measures *adapted from Guillemin F et al. 1993* (Guillemin Francis, Bombardier Claire, Beaton Dorcas 1998).

1.	Translation	Produce several translations Use qualified translators
2.	Back-translation	Produce as many back-translations as translations Use appropriate back-translators but not the same persons as in step 1.
3.	Committee review	Constitute a multidisciplinary committee to compare source and final versions Use structured techniques to resolve discrepancies. Ensure that the translation is fully comprehensible and easily understandable
4.	Pre testing	Check for equivalence in source and final versions using a probe technique or submitting to bilingual people
5.	Weighting of scores	If the answers are given in scores consider adapting the weights of scores to the cultural context

Whether validity, reliability, and sensitivity to change should also be considered in the cross-cultural adaptation process is a matter of controversy. While it is a question in cross cultural adaptation procedure, it is a must for newly developed questionnaires. The VALIDITY analysis answers the questions does this questionnaire measures what we think it does? Out of number of validity measures (content validity, criterion validity, construct validity), the construct validity is mostly important and performed by using the factorial analysis. The factorial analysis gives us a chance to group the questions into subgroups and the questionnaires are usually better answered when they are grouped in a logical manner.

RELIABILITY means how strong can I rely on the results I get from the studied group using the questionnaire or how consistent are the results? Reliability analysis allows you to study the properties of measurement scales and the items that compose the scales (Huck 2004). The Reliability analysis procedure calculates a number of commonly used measures of scale reliability and also provides information about the relationships between individual items in the scale. The measure for reliability analyses are the following:

- test-retest reliability;
- scale internal consistency & item internal consistency,
- inter rater reliability.

Test retest method is intended for measuring stability in time and should show no change in answers when the questionnaire is given repeatedly in a short period of time. The interval of time may be as short as one day or it can be as long as a year or even more. Regardless of the length of time between the two testings, the researcher will simply correlate the two sets of scores by the mean of correlation coefficient which is simply renamed into test-retest reliability coefficient.

A commonly used statistical method for *scale internal consistency* is alpha analysis (Cronbach model). This is a model measuring the *internal consistency*, based on the average inter-item correlation. It is expected to be higher than 0.7. *Item internal* consistency measures how one question in a scale

correlates with the rest of a scale. The second popular procedure on internal consistency involves splitting each examinee's performance into two halves, usually by determining how the examinee did on the odd-numbered items grouped together and the even-numbered items grouped together. Those two halves are then correlated and obtained r is inserted into special formula (called Spearman-Brown) and the final numerical result is called the split-half reliability coefficient.

Inter rater reliability method should be applied when two raters are supposed to give similar answers on the same situation. Inter-rater reliability or concordance is the degree of agreement among raters of a test result.

To quantify the degree of consistency among the raters four popular procedures for doing this are Cohen's kappa, Kendall's coefficient of concordance, the intraclass correlation, and Pearson's product moment correlation. The kappa statistic is the most widely used inter rater reliability coefficient when the variables are nominal or categorical. Generally Kappa coefficient > 0.40 is acceptable. Kendall's procedure is appropriate for situations where each rater is asked to rank the things being evaluated. The third procedure for assessing inter rater reliability to be considered here is called intra class correlation. Abbreviated as ICC, is typically used to estimate the reliability of ratings. For example, each of 20 job applicants might be rated by each of five members of a hiring team. After analyzing the set of ratings, ICC could be used to estimate the reliability, individual rating provided by a single rater or the mean rating provided by a group of raters.

Some of the questionnaires are expected to be able to reflect the changing situation during the time, so it is when they are used for health measures changes before and after situation, for example, surgical intervention, rehabilitation or medical treatment. The changes in scores are considered as SENSITIVITY TO CHANGE analysis. Period of time should be appropriate for changes to occur, for example, recovery period after surgical treatment or treatment effect after drug prescription.

The detailed analysis in cross cultural adaptation and psychometric analysis is presented in the document *Recommendations for the Cross-cultural Adaptation of Health Status Measures* prepared by Beaton D, Bombardier C, Guillemin Fr, Ferraz MB in 1998.

Chapter 6. Cohort studies & prognosis

Cohort study design
How to perform the cohort study
Advantages and disadvantages of cohort studies
How to perform a prognostic study

Cohort study design

Cohort studies are the best method to determine the incidence and natural history of a disease or condition. The starting point of a cohort study is the recording of healthy subjects with and without exposure to the candidate risk factors or the characteristic being studied. The key feature of a cohort study is that subjects are grouped on the basis of their exposure characteristics being though being healthy. All individuals exposed to the agent under study are followed the same time period and usually long enough. The measure of disease or condition incidence (in other word - risk) in both groups of individuals, with or without exposure, is counted and compared. The measure of effect – relative risk is used most often. The directionality of a cohort study is always forward, but the timing may be prospective or retrospective depending on the time when the study was started, prior to the outcome or after the outcome occurred. Most often cohort studies are understood as looking forward by directionality and carried out prospectively in time (Figure 6.1), although it is not always so. It depends on what data are available for the study (Levin 2006b).

Key words:
cohort studies
outcome
prospective cohort
retrospective cohort
follow up period
incidence/risk
relative risk
the Sydney beach users study
prognosis
prognostic factors
inception cohort
survival analysis
censored data

How to perform a cohort study

1. Clearly define the study question you intend to answer.
2. Be sure what outcome, variables and individuals will be studied. For this operational definition, categorization and dummy coding of variables should be defined beforehand.
3. What cohort will be studied: prospective or retrospective
4. For how long the individuals should be followed up if the cohort is prospective.
5. Analyze the data by defining the incidence/risk to develop the outcome and compare the risk between the two groups within the cohort.

The *study question* is to explore the incidence/risk of the disease/outcome and to compare the risk to develop the outcome between the group with and without exposure.

It is always more convenient to have clearly stated **outcome** by definition (for example, myocardial infarction, fracture, rheumatoid arthritis, etc.), though it is not very simple. There are a lot of outcomes which occur silently without pronounced clinical manifestation and the start of disease is not known, like osteoporosis, kidney diseases, some forms of dementia, etc.

For **prospective cohort** a group of people is chosen who do not have the outcome variables of interest (e.g. myocardial infarction) but are at risk to develop it (e.g. young children should not be included in the study for myocardial infarction development since they are hardly at risk to develop myocardial infarction). The investigator then measures a variety of variables that might be relevant to the development of the condition. Over a period of time the individuals in the sample are observed to see whether they develop the outcome of interest (that is myocardial infarction). Those individuals who do not develop the outcome of interest are used as non cases. The **retrospective cohort** studies use databases already constructed for other purposes as well as data collected for practical or clinical purposes. If the data are broad enough and acquired in a trustful manner, then they may be used for retrospective cohort study. The methodology is the same but the study is performed posthoc. The cohort is "followed up" retrospectively (Figure 6.2). The study period may be many years but the time to complete the study is only as long as it takes to collate and analyze the data.

The **follow up period** for prospective cohort or study period for retrospective cohort should be chosen long enough to develop the outcome, usually several years, although it may be several days if the outcome is expected during the forthcoming week.

Analyzing the data in cohort studies means first of all counting the measure of disease, namely, **incidence/risk** of the outcome and measure of effect, so called **relative risk**. The measure of disease in cohort studies is the incidence proportion or incidence rate, which is proportion of subjects who develop the disease under study within a specified time period. The numerator of the rate is the number of diseased subjects and the denominator is usually persons if they were followed for approximately the same time or the number of persons-years of observation. The incidence rates for exposed and non-exposed subjects are calculated separately. The measure of effect between exposure and disease in cohort studies is the relative risk. The relative risk is the ratio of the incidence rate among exposed individuals to that of control subjects. Measures of disease frequency and effect are approached in depth in the first lecture of this book.

Let us follow the classical cohort study performed in Sydney (Australia) and published in 1993(Corbett et al. 1993). **The Sydney Beach Users Study** is an example of application of epidemiologic principles and methods to investigate a localized public health issue. The background for this study started from complains that the popular beaches surrounding the city were becoming more and more unsafe for swimming because of the pollution. The *primary research question* of interest was: are persons who swim at Sydney beaches at increased risk for developing an acute

infectious illness. The study was carried out by selecting subjects on the beaches throughout the summer month of 1989–90 (December–February). Those individuals eligible to participate at the initial interview at the beach were then followed up by phone a week later to determine swimming exposure on the day of the beach interview and subsequent illness status during the week following the interview.

In this study the health outcome variable of interest was whether or not a person swimming at a beach develops an acute infection such as cough, cold, flu, ear infection, gastrointestinal or eye infection, within one week of swimming at the beach. The study team decided to use self-reported symptoms of illness obtained by telephone interview of study subjects 7 to 10 days after the initial interview. The most important exposure variable according to the investigators was swimming status on the day of interview. They defined swimming as any immersion of the face and head in the water. It was decided that the persons self-reporting of swimming was the only feasible way to obtain swimming information. Subjects were excluded from the study if they reported swimming in the previous 5 days or having an illness that prevented them from swimming. Subjects were included if they were at least 15 years old and agreed to both an initial beach interview and a follow up telephone interview. A complex sample survey design was used to obtain the nearly 3000 study participants. (Note, for cohort studies larger sample size is needed than for case control studies). Since the participants were followed up prospectively and forward for a time period of 7 days, the appropriate study design was prospective cohort. Once the study design has been determined, appropriate measures of disease frequency and effect can be specified. A measure of disease frequency provides quantitative information about how often the infectious disease occurred in the total sample of individuals, i.e. what the incidence or risk to contract the infectious disease is if attending the beach. Similarly, if we want to measure the incidence of illness in two different groups we divide the number of those who contracted the disease in two different groups by the number of swimmers and non-swimmers.

The information for analyzing the data can be described in the form of a two way table which is not very much liked by researchers when computer analysis can be performed in two minutes but is very useful to get a feeling of the data and to check them (Table 6.1).

Table 6.1. 2x2 tables (A; B) for counting measures of frequency and association in cohort studies

A

		Disease		Total
		Yes	No	
Swim	Yes	a	b	a+b
	No	c	d	c+b
Total		(a+c)	(b+d)	a+b+c+d

B

		Disease		Total
		Yes	No	
Swim	Yes	532	1392	1924
	No	151	764	915
	Total	683	2156	2839

In result, incidence or risk to develop an infectious disease during the period of ten days if attending the beach was 683/2839 = 0.24 or 24%

The risk to develop a disease among swimmers was 532/1924 = 0.277 or 27.7%

The risk to develop a disease among non-swimmer was 151/915 = 0.165 or 16.5%

Referring to lecture 1, if we want to compare two measures of disease frequency, such as two risks, we can divide one risk by the other.

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{R_1}{R_0}$$

This measure of effect or association is called a risk ratio (RR): 27.7% / 16.7% = 1.68. This means that swimmers have a risk for the illness that is 1.68 times higher the risk for non-swimmers (Figure 6.3). Do not get into problem when trying to learn the formula by heart. This fits well only when the disease is displayed in columns and exposure in rows otherwise it would be wrong and would require the following formula:

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

The following example will illustrate the difference between risk ratio and odds ratio in numbers.

The table below (Table 6.2) summarizes the results of a five-year follow up study to determine whether or not smokers who have had heart attack will reduce their risk for dying by quitting smoking. A cohort of 156 heart attack patients were studied all of whom were regular smokers up to the time of their heart attack. Seventy five continued to smoke after the attack. The other 81 patients quit smoking during their recovery period. Of 75 patients that continued smoking 27 died. Of 81 who quit smoking 14 died. Here is the scheme for this study:

Table 6.2. The scheme of heart attack cohort study

	Heart attack patients		Total
	Smoke	Quit	
Death	27	14	41
Survival	48	67	115
Total	75	81	156

$$RR = \frac{a/(a+c)}{b/(b+d)} = \frac{27/75}{14/81} = \frac{0.36}{0.17} = 2.1$$

$$OR = \frac{ad}{bc} = \frac{27 \times 67}{14 \times 48} = 2.7$$

The fact that these two numbers (the risk ratio and odds ratio) are not equal should not be surprising, since the risk ratio and odds ratio are two different measures. But the values of both of them are not very different. They both suggest that there is a moderate relationship between quiting smoking and survival status. Note, while you can estimate RR and OR for cohort studies, doing the same for case control studies would be a mistake. Primarily because of study

groups were assembled in different ways. In cohort studies the groups of patients are assembled according to the exposure and it allows approximations to risk evaluation. In case control group the group is assembled according to the available outcome cases and the control group is artificially collected what gives no opportunities to measure the risk and the best approximation to the risk is OR in case control studies.

Advantages and disadvantages of cohort studies

The research on risk factors heavily lies on cohort studies as no other studies are strong enough to estimate risk. Moreover some risk factors may appear to be not only risk but also the etiological cause of the disease. And there is no other way to test for causation but to perform a cohort study. This is the case with the classical study of R. Doll and A.B. Hill conducted in 1947 where they showed in a case control trial that smoking may be related to lung cancer[5]. However, they managed to prove the causational relation between smoking and lung cancer by conducting a cohort study categorizing a group of British physicians according to their smoking histories and then analyzing the causes of death among those who died, to see whether cigarette smokers had the higher incidence of lung cancer (Doll et al. 2004).

A further advantage is that a single study can examine various outcome variables. For example, cohort studies of smokers can simultaneously look at cardiovascular and cerebrovascular disease in the same cohort. This contrasts with case control studies as those assess only one outcome variable. Cohort also permits calculation of the incidence rates and effect of each variable on the probability on developing the outcome of interest, namely relative risk. There is no other way to learn about the incidence of a disease/condition unless a cohort study is performed and the individual is followed up for substantial period of time. To add to the advantages, both the methodology of performance as well as performance are easily understood by every researcher.

Alongside with advantages, disadvantages are present. However, where a certain outcome is rare then a prospective cohort study is inefficient. The efficiency of a prospective cohort study increases as the incidence of any particular outcome increases. Another problem with prospective cohort studies is the loss of some subjects to follow up. This can significantly affect the outcome. Taking incidence analysis, the loss of a few cases will seriously affect the numerator and hence the calculated incidence. The rarer the condition, the more significant is the effect.

The cohort studies are expensive and especially the prospective cohorts where the data are collected prospectively according to a flow sheet and the information on individual status are updated periodically. Variables that are relatively fixed, for example height, need only be recorded once. Where change is more probable, for example, change of medication, variations in blood pressure or laboratory, repeated measurements are required more often. Retrospective studies are much cheaper as the data have already been collected. However, because the cohort was originally constructed for

5   http://en.wikipedia.org/wiki/British_Doctors_Study

another purpose it is unlikely that all the relevant information will have been scrupulously collected (information bias). Another source of bias is that the cohort studied is not representative of the population it was drawn from. For example, individuals interviewed in the teaching hospital do not represent the individuals who are admitted to peripheral hospitals, or employed people do not represent population at large, as employment is itself associated with generally better health than unemployed people (selection bias). Similarly people who respond to questionnaires tend to be more motivated to do so than those who do not, so the latter are omitted.

The major disadvantage of cohort studies that is even more pronounced than in a case control studies is the inability to control all other factors that might differ between the two groups, so called confounding variables. The only way to eliminate all possibility of a confounding variable is via a prospective randomized controlled study, which is rarely feasible in public health studies. To minimize the influence of confounding variables means to obtain the information on all probable relevant confounders, because control on confounding variables to certain level is possible through statistical analyze.

To summarize, the main advantages of cohort studies are:
- Permit calculation of incidence proportion and rate as well as relative risk
- Provide information on etiology of disease not only risk
- The information on exposure variable can be *ad maximum* detailed
- Give an opportunity to study multiple outcomes as well as multiple exposures
- Methodology and calculations are easily understandable and simulate the real life situation and the results from this kind of studies are anticipated by community

Disadvantages of cohort studies are:
- Not suited for studying rare diseases because a large number of subjects is needed
- They are expensive because the large sample size is needed and periodical interviews are required
- The drop outs during the study are usual and minimizing them is difficult

---

The cohort studies described above are aiming to describe the risk to develop disease/ condition during the time period. They start with assembling the healthy individuals and are sometimes called the cohort studies of risk. Similar to cohort studies of risk there are studies of **prognosis**. While cohort studies of risk usually deal with healthy people, the studies of prognosis deal with the outcomes of sick people. Conditions that are associated with an outcome of sick people disease are called **prognostic factors**. Prognostic factors are analogous to risk factors, except they represent a different part of a disease spectrum; from disease to its outcome. Risk and prognosis describe different phenomena though sometimes they are used interchangeably. For risk, the event being counted is the onset of disease. For prognosis, a range of disease consequences is counted,

including death, complications, disability and sufferings. Variables associated with an increased risk are not necessarily the same as those marking a worse prognosis and sometimes they are considerably different for a given disease. For example, low blood pressure decreases the chances of having an acute myocardial infarction, but it is a bad prognostic factor for a patient who has already developed acute myocardial infarction. Risk and prognostic factors for acute myocardial infarction are listed in the Figure 6.4 adapted from *Clinical Epidemiology by R Fletcher & S.Fletcher* (Fletcher, Fletcher 2005a). Cohort studies of prognosis consider the clinical course of disease if the disease was treated or natural course of disease if no intervention was applied.

How to perform a prognostic study

The performance of prognostic studies meets the same requirements like cohort studies of risk factors. After having clearly stated what *outcome*, namely prognosis, of what cohort of patients is going to be described, the appropriate sample size of patients is needed. At best, studies of prognosis include a complete description of patients and the setting in which they were identified. If the national registries for certain diseases are available, one can be confident that the patients randomly selected from this database to the study can form an unbiased *sample* of all such patients. But it is rarely so. Most studies of prognosis, however, are based on clinical samples not directly related to geographical populations and it is not always that the results from clinical samples can be generalized to all the patients having the particular disease. They begin observation at a specified *point in time*, usually at the beginning of disease. If the patients are assembled in the beginning of the disease the term used is – **inception cohort**. It is absolutely needed to assemble patients at the same point in time, be it onset of symptoms, time of diagnosis or time of beginning of the treatment. The follow up should be adequate and long enough for clinically important outcome events to occur.

Description of prognosis may include the full range of disease complications and manifestations, along with recovery and remission. Nevertheless the estimates of a bad prognosis, for example death, are referred more often in scientific literature than other ones.

It is convenient to summarize the course of disease as a single rate – a proportion of people experiencing an event in a fixed time period. Some of them are delineated in the Table 6.3. These rates have in common the basic incidence components: events arising in a cohort of patients over time. Risk ratio for prognostic studies is named hazard ratio. Computing hazard ratio requires more complicated statistics while the interpretation is quite similar to relative risk.

Rates can be easily remembered and interpreted, they are useful for comparisons but their drawback is that they lack the information at a particular point in time. To describe the prognosis at the set point in time and to make the use of all available data from each patient in the cohort, regardless to the follow up period, the **survival analysis** was developed. It has got a wide application in clinical trials and is becoming more and more popular among researchers. The usual method

Table 6.3. Measures commonly used to describe prognosis

		Definition
Rate	5 – year survival	Proportion of patients surviving 5 years from some point in the course of disease; can be counted only when time of follow up is long enough
	Case fatality	A proportion of people with a particular disease who die of it
	Disease specific mortality	Number of individuals dying of a specific disease expressed per population of 100 or 10^3....10^6
	All-cause mortality risk	Number of individuals dying of a whatever cause expressed per population of 100 or 10^3....10^6
	Response	Percent of patients showing some evidence of improvement following an intervention
	Remission	Percent of patients entering a phase in which disease in no longer detectable
	Recurrence	Percent of patients who have return of disease after a disease-free interval
Ratio	Hazard ratio	Analogous to the risk ratio, hazard ratio (HR) is used to compare the chance of event happening in two or more groups. If HR is 1.0 the two groups have the same chance of having the same event. Statistically HR is derived from survival analyses and cannot be counted as simply as risk ratio.
Survival analyses	Survival curves	Cumulative proportion of surviving patients during the follow up period. When an event other than survival is described, the general term is time-to-event analysis.

named after its authors is called Kaplan-Meier analyses. Survival analysis can be applied to any outcomes that are dichotomous and occur only once during follow-up (like time to myocardial infarction or time to recurrence of cancer).

For those who are interested in survival analysis in more depth we suggest following the data of ten patients presented in Table 6.4 and referring to the survival time in years following treatment for malignant melanoma of the skin.

Table 6.4. Calculation of Kaplan-Meier estimate of the survivor function (adapted from series What is?)[6]

A. Survival time (years)	B. Number still in the study at start of time	C. Number of death	D. Number censored	E. Proportion surviving until end of interval	F. Cumulative proportion surviving
0.9	10	1	0	1 − 1/10	0.90
1.1	9	1	0	1 − 1/9	0.80
1.3*	8	0	1	1 − 0/8 = 1	0.80
1.3	7	1	0	1 − 1/7	0.69
1.5	6	1	0	1 − 1/6	0.57
2.7	5	1	0	1 − 1/5	0.46
2.7*	4	0	1	1 − 0/4 = 1	0.46
2.7	3	1	0	1 − 1/3	0.31
3.5*	2	0	1	1 − 0/2 = 1	0.31
4.1*	1	0	1	1 − 0/1 = 1	0.31
*indicates a censored survival time					

To determine the Kaplan-Meier estimate of the survivor function for the above example, a series of time intervals is formed. Each of these intervals is constructed to be such that one observed death is contained in the interval, and the time of this death is taken to occur at the start of the interval.

Table 6.4 shows the survival times arranged in ascending order from shorter survival time to the longest. Some survival times are **censored data**, meaning that these patients have withdrawn from the study for other reasons or survived the whole follow up period of the study. In other words, those who did not experience an outcome are censored.

The number of patients who are alive just before 0.9 years is ten and they are showed in column B. Since one patient dies at 0.9 years (column C), the probability of dying by 0.9 years is 1/10 = 0.10. So the corresponding probability of surviving up to 0.9 years is 1 minus the probability of dying (column F) or 0.900. The cumulative probability of surviving up to 1.1 years, then, is the probability of surviving at 1.1 years, and surviving throughout the preceding time interval – that is, 0.900 x 0.889 = 0.800 (column F). The third time interval (1.3 years) contains censored data, so the probability of surviving is unchanged from the previous interval. This is the Kaplan-Meier estimate of survival function. The computer based programs generate the curve similar to the described table above (Figure 6.5). A plot of the Kaplan Meier estimate of the survival function is a step function, in which the estimated survival probabilities are constant between neighboring death times and

only decrease at each death. We can also interpret the curve by explaining the proportions. For example, almost all of the cohort live the entire first year after the surgery but only third of them remain alive for four years after surgery. The median survival time can also be calculated from the table. It is the time at which half the individuals have reached the event of interest (death in this case). If the survival curve does not fall to 0.5 (50%) then the median time cannot be computed. For this cohort a median survival time after surgery due to melanoma is 2.7 years. We can generalize these results to the population of patients for whom this sample is representative.

To conclude, follow up studies are difficult to perform but there is no other way in biomedicine or public health to get to valid and trustful results.

Chapter 7. Screening & Diagnostic test studies

Screening studies
Diagnostic test studies

Screening studies

This lecture introduces the reader to some terms like, sensitivity, specificity, etc. which are commonly used in daily practice but frequently one cannot tell what they stand for. There are two settings, namely, a healthy population where the screening tests may be used and a clinical setting where diagnostic test studies are performed. Sensitivity, specificity, as well as predictive values are the measures of test accuracy, be it for screening or diagnostic purposes. Mathematically the results for the screening programs and diagnostic tests are calculated the same way.

Screening is the identification of unrecognized disease or risk factor by history taking (e.g. asking if the patient smokes), physical examination (e.g. measuring the blood pressure), laboratory tests (e.g. faecal occult blood test) or instrumental examination (mammography). Most screening tests are performed in clinical settings, usually at GP level. They are performed at the state health budget cost, so their benefit and cost balance should be evaluated beforehand. Members of the general public are typically invited to undergo screening tests of various sorts which enables to separate them into those with higher and lower probabilities of disease. Those with positive tests or higher probabilities are then urged to seek medical attention for definitive diagnosis. Those with negative test or lower probabilities receive no direct health benefit because they most probably do not have the disease condition being screened.

A good screening test must, therefore, have a high sensitivity, so that it does not miss the few cases of disease that are present, the so called false negative cases, and a high specificity, to reduce the number of people with false-positive results who do not have a disease (Hennekens, Buring & Mayrent 1987). Sensitivity and specificity are determined for screening tests very much the same as they are for diagnostic tests, except that the gold standard for the presence of disease usually is not another test but rather a period of follow up. Following up the classical study about the screening for colorectal character, the sensitivity was counted by detection method. The ratio of the number of colorectal cancers detected by screening for occult blood (confirmed histologically with a follow up colonoscopy and biopsy) was divided by the number of colorectal cancers detected by

Key words:
Screening
Sensitivity
Specificity
Feasibility
Positive predictive value
Negative predictive value
Effectiveness
Diagnostic studies
Likelihood ratios
ROC curves

screening plus the number discovered during the year after each screen (Fletcher, Fletcher 2005b). The assumption was made that any cancer diagnosed during the year following the negative screen was an interval cancer present at screening but missed, and the result was false negative. In the study, 178 cancers were detected at the time of screening, and 21 cancers became evident during the subsequent year in people who tested negative. Thus sensitivity was calculated as 178 divided by (178 + 21), or 89.4%.

Another classical study for breast cancer screening illustrates the way how the **sensitivity** and **specificity** of the screening test can be calculated if randomized preventive trial can be applied (Shapiro, Goldberg & Hutchison 1974). A total of 62,000 women New York aged 40 to 64 years were identified for the study. The women were offered an initial screening examination for breast cancer (consisting of a combination of mammography and a physical examination) followed by three follow up examinations at yearly intervals.

As shown in the Table 7.1, a total of 64,810 screening examinations were performed among the study population.

Table 7.1. Calculations of sensitivity and specificity of breast cancer screening examination

Screening test (physical examination and mammography)	Breast cancer (confirmed by biopsy)		Total
	Cancer confirmed N=177	Cancer not confirmed N=64,633	
Positive	132	983	1,115
Negative	45	63,650	63,695
Total	177	64,633	64,810

During the first 5 years of observation, 132 breast cancers were diagnosed among 1,115 biopsies that were recommended on the basis of the results of the screening procedures. In addition 45 cases of breast cancer were detected among women who screened negative but were diagnosed with the disease during the subsequent years. These women were assumed to have been false negatives, that is, they were assumed to have had a disease at the time of the screening test. Thus the sensitivity of mammography plus physical examination in these data would be 132/177 or 74.6%. This means that of those diagnosed with breast cancer during the study period, approximately 75% tested positive on the screening procedure. The specificity of the screening program would equal 63,650/64,633 or 93.5%, indicating that virtually all women who did not have the disease tested negative.

$$\text{Sensitivity} = \frac{a}{a+c} = \frac{132}{177} = 74.6\%$$

$$\text{Specificity} = \frac{d}{b+d} = \frac{63,650}{64,633} = 98.5\%$$

Even after the disease is determined to be appropriate for screening and a valid test becomes

available, it remains unclear whether a widespread screening program for that disease should be implemented. Evaluation of a potential screening program involves consideration of two issues: first whether the proposed program is feasible, and, second, whether it is effective.

The **feasibility** of a screening program is determined by a number of factors related to program performance, including cost effectiveness, acceptability, and the potential to yield the cases. An ideal screening test should be simple and low cost. It should take only a few minutes to perform, require minimum preparation by the patient, depend on no special appointment, and be inexpensive. Simple and quick examinations such as blood pressure determinations are ideal screening tests. Conversely, complicated diagnostic tests such as colonoscopy, which are expensive and require bowel preparation, are reasonable in patients with symptoms and clinical indications but may be unacceptable as screening tests, especially if they must be repeated frequently.

The screening test should be also safe. It is reasonable to accept certain risk with diagnostic test for people who have complains and it is acceptable in clinical practice. Without the diagnostic test there would be no diagnosis. It is quite another matter to subject people to some risk when they do not have complaints. Since screening test is associated with discomfort, it is not anticipated by healthy population first, and patients as well. There is also a clinician's factor of performing screening tests. The clinician's acceptance is especially relevant for screening tests that involve clinical skill, such as mammography or colonoscopy. Performing a procedure repeatedly can be tiring when the vast majority of results are normal. Adverse effects of screening test include not only discomfort during the test procedure and possible complications but also false positive test results and over-diagnosing and false negatives results and misdiagnosing also. A false positive screening test result is an abnormal result in a person without a disease. Some tests, especially those which aim to predict a disease of low prevalence may generate false positive results quite often. The false positive results account for only a small minority of screening test results (only about 10% of screening mammograms are false positive), but even so they can affect a large percentage of people who get screened. False negative results are normal results found in individuals who have a disease but have no symptoms yet. They generate not from physicians mistakes but from the test ability to find the abnormalities, e.g. the test for occult blood does not "catch" the bowel cancer which is small enough to produce even minor bleeding.

With respect to the yield of cases, or number of cases detected by a screening program, one commonly considered measure is the predictive value of the screening test. The predictive value measures whether or not an individual actually has the disease. **Positive predictive value** PV(+) is the probability that a person actually has the disease, given that he or she tests positive and is calculated as: $PV(+) = \dfrac{a}{a+b}$. Analogously, **negative predictive value** PV(−) is the probability that an individual is truly disease – free given a negative screening test and is calculated as follows: $PV(−) = \dfrac{d}{c+d}$. The calculation of these measures can be illustrated using data from Breast Cancer

Screening Project presented earlier in this lecture. A positive predictive value, or the probability that the woman who tested positive on the screen actually had breast cancer, is 132/1115 or 11.8%. This indicates that approximately 1 out of every 8 women who were referred for diagnostic evaluation after testing positive on mammography and physical examination actually did have breast cancer. Negative predictive value, or the probability that the woman who tested negative actually did not have breast cancer is 63,650/63,695 or 99.9 %. Thus, virtually all women who tested negative were in fact free from the disease. The positive predictive value of a screening test can be increased by increasing the prevalence of preclinical disease in the screened population. Thus screening for breast cancer, targeting the program to women with a positive family history will detect more cases than screening among general population because the prevalence of preclinical disease in that setting is expected to be higher.

The second, and ultimately most important, aspect of evaluation of a screening program is whether it is effective in reducing morbidity and mortality from the disease. The most definitive measure of **effectiveness** is the comparison of cause specific mortality rates among those whose disease was picked up by screening and those whose diagnose was related to the development of the symptoms. After 9 years of follow-up, there was an overall statistically significant reduction in breast cancer mortality among women who were offered screening compared with women randomly assigned to receive their usual medical care in Breast Cancer Screening project (Shapiro, Goldberg & Hutchison 1974). Simple observation for incidence or mortality cases among population may be useful to prove the effectiveness of the screening program if the reduction for incident cases or fatal cases can be proved.

Following international practice, six preventive or screening programs are on-going in Lithuania in the meantime.

The screening of a population with a high cardiovascular risk since 2005
Cervical cancer screening programme since 2004
Screening for prostate cancer since 2006
The mammography screening programme since 2005
Children dental programme to reduce dental decay since 2005
Screening for colorectal cancer since 2009

Diagnostic test studies

While the idea of screening programs are to prevent the individual from disease development and to predict the disease in a preclinical stage, the idea of a diagnostic test is to diagnose the disease and preferably in early stages and at low cost and in convenient way for both patient and physician. A diagnostic test is any kind of medical test performed to aid in the diagnosis or detection of a

suspected disease or condition. This is different from a screening test which is used when not a disease or condition is suspected, but when people are considered to be at high risk of developing a disease or condition.

A diagnostic test is ordinarily understood to mean a test performed in a laboratory, but the principles discussed in this chapter apply equally well to clinical information obtained from history, physical examination, and imaging procedures. In clinical medicine, studies concerned with diagnostic tests are usually called **diagnostic** test **studies**. The purpose of a diagnostic study is to evaluate the new test performance in relation to the so called golden standard, somewhat of knowing whether the disease is truly present or not, or simply the test that is usually used to diagnose the particular disease. Typically, the "gold" standard is a test that is more detailed, expensive, or risky than the diagnostic test used by the physician and the latter is supposed to be evaluated for diagnosing the disease.

Table 7.2. The relationship between a diagnostic test result and the true disease in a clasification table

TEST		DISEASE	
		Present	Absent
	Positive	True positive a	False positive b
	Negative	False negative c	True negative d

A simple way of looking at the relationships between a new test results and the true diagnosis set by a golden standard is shown in the classification table (Table 7.2) which is indeed 2x2 table discussed above.

Imagine a hypothetical situation that the condition is a peripheral vascular disease (PVD). In a diagnostic study for PVD, the clinicians target only patients with a specific symptom, for example "pain in the leg" and then performs both diagnostic test, typically ultrasound, and the golden standard procedure, typically in this case considered, computerized angiography. Here are the results of such a diagnostic test study in the form of a classification table (Table 7.3).

Table 7.3. Table for calculation of diagnostic values of diagnostic test for peripheral vascular disease confirmation

Diagnostic test (ultrasound examination)	Peripheral artery disease	
	Confirmed N=60	Not confirmed N=140
Positive	48	14
Negative	12	126
Total	60	140

Using the diagnostic test, namely ultrasound, the test results are labeled as positive (+) or negative (–) on the rows of the table. The results using the gold standard, angiography, are labeled on the columns on the table. Using the information in the classification table, the accuracy of a diagnostic test can be evaluated using several important measures, including sensitivity, specificity and predictive values.

The sensitivity describes the test's performance in patients who truly have a disease, and is defined as the probability of a true positive test result in present disease column: a/(a+c). In the given example it is 48/60 = 0.80. Specificity describes the test performance among patients who are truly without the disease. It is defined as the conditional probability of a negative test result in the absence of disease: d/(b+d). In the given example it is 126/140 = 0.90. Both the sensitivity and specificity are popular in common language, but the predictive values yield more information in clinical setting, to predict what is the probability for the patient to have the disease or to have not.

Similary to the screening studies, there are two possible results for a test, so there are two different predictive values. The probability of actually having the disease when the test is positive is called the positive predictive value (PV(+)), the probability of actually not having the disease if the test is negative is the negative predictive value (PV(–)). The positive predictive value is calculated as the number of true positive results divided by all positive results: a/(a+b); 48/62 = 0.77 in a given example. The positive predictive value is often referred to as the post-test probability of having a disease. The negative predictive value is the number of true negative results divided by the total number of subjects with negative test results: d/(c+d); 126/138 = 0.91. The closer these proportions are to 1, the better the test's accuracy. It should be noted that the predictive values greatly depend on the prevalence of the disease in the particular setting they were counted. The prevalence of true disease can greatly influence the size of the predictive values above. In the given example the prevalence of PVD is 60/200 = 0.3, meaning that the values above work best in the setting with the disease prevalence around 30%. In other words, the ultrasound shows positive and negative predictive values of 0.77 & 0.91, respectively, in a setting where there are a lot of patients with PVD, namely angiology department. In outpatient clinics where the patients with DVT are seen rarely, it may show much less accuracy than in the previous example. Building up the classification table for a certain disease should always begin with counting the prevalence to get the feeling as to what kind of clinical setting the results can be applied.

Likelihood ratios (LR) are an alternative way of describing the performance of a diagnostic test. They summarize the same kind of information as sensitivity and specificity and can be used to calculate the probability of a disease after a positive or negative test (positive or negative predictive value). An advantage of likelihood ratios is that they can be used at multiple levels of test results. Earlier described method for test accuracy is possible when dichotomous data are available as a diagnostic test result. When the diagnostic test result is a continuous data, then the method called receiver operator characteristic (ROC) curve is a better way to decide which of the values of the diagnostic test displayed on a continuous scale shows the best sensitivity and specificity. ROC curve method takes them both into one common measure – likelihood ratios. The LR for a particular value of a diagnostic test is defined as the probability of that test result in people with the disease divided by the probability of the result in people without disease. LR express how many times more (or less) likely a test result is to be found in diseased, compared to non diseased, people. In the case of a test's positive LR(+), it is the ratio of the proportion of diseased people with a positive test result

(sensitivity) to the proportion of non diseased people with a positive result (1-specificity). A test's negative LR(–) is calculated when the test result is negative. In that case, it is the proportion of diseased people with negative test results (1-sensitivity) divided by the proportion of non diseased people with a negative test result (specificity). A value of 10 or more of LR(+) and 0.1 or less of LR(–) indicates that the test is extremely good.

We present the data of 20 patients for whom the diagnostic test – antibodies against citrulinated proteins (A-CCP) was applied to evaluate their diagnostic accuracy (Table 7.4). A-CCP test is a newly introduced test into the clinical practice for rheumatoid arthritis diagnosis.

Table 7.4. The levels of antibodies against citrulinated proteins (A-CCP), diagnostic test of rheumatoid arthritis in a set of 20 patients

Diagnosis of Rheumatoid arthritis	ACCP level (optic units)	Diagnosis of Rheumatoid arthritis	ACCP level (optic units)
No	5	Yes	16
No	6	No	17
No	7	Yes	17
No	8	Yes	18
No	9	No	18
Yes	10	Yes	19
No	12	Yes	20
No	12	Yes	23
No	16	Yes	24
Yes	16	Yes	25

If the cut off level of the test were set too low (9 and lower),the sensitivity is high (100%) but the specificity is low (50 and lower), meaning that many people without rheumatoid arthritis would be misclassified (Table 7.5) as having the disease. On the other hand, if the cut off level were set too high, many patients with rheumatoid arthritis will be missed. The computer based program suggests that the best trade-off between sensitivity and specificity would be 12 optical units. This values yields LR(+) value 3 and LR(–) value 0.14 what is quite good for the laboratory test.

The scheme of the ROC curve illustrates visually the relationship between sensitivity and specificity for a given A-CCP test (Figure 7.1). It is constructed by plotting the true-positive rates (sensitivity) against false positive rate (1-specificity). Test with sufficient characteristics goes up toward the upper left corner of the ROC curve, like it is in the case of A-CCP testing. Tests that perform less well have curves that fall closer to diagonal running from lower left to upper right. Generally the best cut-off point is at or near the "left shoulder" of the ROC curve. The overall accuracy of the test can be described as the area under the ROC curve; the larger the area, the better

Table 7.5. The relation of sensitivity and specificity when using A-CCP levels to diagnose rheumatoid arthritis. Area under curve – 0.880 (significance level – 0.0001)

A-CCP level	Sensitivity	Specificity	LR(+)	LR(-)
≥5	100	0	1.0	
>9	100	50	2.00	0.00
>10	90	50	1.80	0.20
>12*	90	70	3.00	0.14
>16	70	80	3.50	0.38
>17	60	90	6.00	0.44
>18	50	100		0.50
25	0	100		1.00

the test. Area under curve (AUC) is a measure of overall test validity, the bigger AUC, the more valid is the test. Ideal AUC is 1.0 meaning the highest validity and the worst AUC is 0.5. In our example it is 0.88 and significantly differs from area 0.5 covered by diagonal.

Obviously tests that are both sensitive and specific are highly appreciated and can be of enormous value. However, practitioners often work with tests that are neither sensitive nor specific. For true diagnosis the most common way is to use the results of several tests taken together.

Chapter 8. Design and planning of epidemiological studies: the study protocol

Basic principles of an epidemiological study protocol
Requirements for the study protocol

Basic principles of an epidemiological study protocol

In previous chapters, we covered the basic concepts in epidemiology. In this chapter we will concentrate on the principles of how to design an epidemiological study protocol. Only general issues will be covered.

Generally, research is understood to follow a certain structural process. The following steps are usually part of most formal research, both basic and applied:

- **Formation of the topic**
- **Hypothesis**
- **Conceptual definitions**
- **Operational definitions**
- **Gathering of data**
- **Analysis of data**
- **Test, revising of hypothesis**
- **Conclusion, iteration if necessary**

Key words:
topic
hypothesis
conceptual definitions
operational definitions
gathering of data
analysis of data
revising of hypothesis
conclusion

Analytic epidemiological study should also meet these steps.

The study protocol is a document that describes the objective(s), design, methodology, statistical considerations and organization of a study. The protocol should provide a clear description of why the study is being undertaken, the methods to be employed and how the results will be analyzed. It should also address the ethical considerations, data protection procedures, project organization, quality control, time schedule and study diary, publication, and budget.

According to Miettinen OS (Miettinen 1985)(1985) a study protocol should have five purposes (Miettinen 1985): 1) crystallize the project to the researchers themselves; 2) give referees the possibility to review the project (especially for funding); 3) inform and educate all those taking part in the project; 4) ensure the main researchers do not forget any details of the plan in the course of the study; 5) document the procedures of the project for the future.

Requirements for the study protocol

The protocol of the study should be written according to the specifications of the funding body. Although the layout of the application forms varies from one funding body to another, they are generally divided into the following sections (adapted from "Guidelines for Good Pharmacoepidemiology Practices (GPP)" 2007):[7]

A. Descriptive title and version identifier (e.g. date);

B. The names, titles, degrees, addresses, and affiliations of all responsible parties, including the principal investigator, co-investigators, and a list of all collaborating primary institutions and other relevant study sites;

C. The name and address of each sponsor;

D. An abstract of the protocol;

E. The proposed study tasks, milestones, and timeline;

F. A statement of research objectives, specific aims, and rationale;

G. A critical review of the literature to evaluate pertinent information and gaps in knowledge;

H. A description of the research methods;

I. A description of plans for protecting human subjects;

J. A description of plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication;

K. Resources required conducting the study;

L. Bibliographic references;

M. Dated amendments to the protocol.

The title of the Protocol should convey the main purpose (research question) of the research. The title is usually the first part of the protocol to be read and therefore should convey maximum information in a few words. The title should indicate the area of research, introduce the research question and specify the research method to be used. Title page should include the following: title of the research Project, names of the investigators, version number of the protocol, and date of completion of the protocol. The title page should also include the signature of the Principal Investigator.

An abstract of the protocol should be no more than 300 words and at the most a page long (font size 12, single spacing). Provided preferably on a separate page, it should summarize all the central elements of the protocol, for example the rationale, objectives, methods, populations, time frame, and expected outcomes. It should stand on its own, and not refer the reader to points in the project description. (Recommended format for a Research Protocol, WHO http://www.who.int/rpc/research_ethics/format_rp/en/index.html[8])

---

7 http://www.pharmacoepi.org/resources/guidelines_08027.cfm#1)

8 http://www.who.int/rpc/research_ethics/format_rp/en/index.html

The proposed study tasks, milestones, and timeline. The time schedule concerns the sequence and interdependency of different operational tasks and resources. The protocol should specify the time that each phase of the project is likely to take, along with a detailed month by month timeline for each activity to be undertaken.

A statement of research objectives, specific aims, and rationale. The Rationale specifies the reasons for conducting the research in light of current knowledge. It should 'justify' the need for the proposed study by clearly indicating its originality and the potential significance of its findings. It summarizes what is already known of the problem, what is not known and whether there are conflicting views. It is the equivalent to the introduction in a research paper and it puts the proposal in context. This section should include a clear explanation of the main research question i.e. the hypothesis to be tested.

The rationale should not be an exhaustive literature review. At the end the reader should have a clear idea of what is the research question, an understanding that it is original and relevant, and how this research will help fill the gap in the literature.

The research objectives of the study should be defined. It is prudent not to list too many objectives. No more than two or three primary objectives should be defined. If necessary, these can be divided into secondary objectives. Primary objectives must be achieved. They dictate design, sample size and methods. Secondary objectives are of interest, but not essential. In defining secondary objectives, consideration could be given to time and cost, which may impose constraints and choices, for example in terms of sample size, duration of follow-up or data collection

If relevant include a clearly defined hypothesis here. The research question might be purely descriptive, exploratory or explanatory (hypothesis driven). Therefore, not all studies need a clear hypothesis.

 A hypothesis is a proposed explanation for a phenomenon. Hypotheses flow from the problem statement, literature review, and theoretical framework. A hypothesis attempts to answer the question posed by research problem. It is a vehicle for testing the validity of the theoretical framework's assumptions. A hypothesis is a bridge between theory and the real world. Researchers derive hypotheses from theories and subject the hypotheses to empirical testing. Therefore, research objectives should be translated into hypotheses that can be tested statistically. An epidemiological hypothesis is a testable statement of a putative relationship between an exposure and disease. A typical epidemiologic research question describes the relationship between a health outcome variable and an exposure variable taking into account the effects of other variables already known to predict the outcome. "An epidemiologic hypothesis specifies: 1) the characteristics of persons in the population to which the hypothesis applies; 2) the exposure being considered and its interdependencies with other disease determinants; 3) the expected effect of the exposure on disease occurrence; 4) the expected change in incidence associated with a given exposure dose (dose-response relation); 5) the time period that will elapse between the exposure and its putative effects (time-response relation);

6) the extraneous factors that will be controlled during analysis" (Gerstman 2003). Therefore, the hypothesis must be fully elaborated. The hypothesis should be clear, testable or resolvable, state the relationship between exposure and disease, limited in scope, not inconsistent with known facts, supported by literature, theory, references. Hypotheses should be formulated before the study, because they will provide direction for the data collection, analysis and interpretation.

A critical review of the literature. A critical literature review is a critical assessment of the relevant literature. Critical appraisal is the process of carefully and systematically examining research to judge its trustworthiness and its value and relevance in a particular context. There are several questions which should be asked of all research papers, irrespective of the method which has been used:

- Are the aims clearly stated?
- Is the design appropriate to the aims?
- Is the study population defined appropriately?
- How is the sample size justified?
- Are the measurements likely to be valid and reliable?
- Were databases built properly?
- Are the statistical methods described?
- Did untoward events occur during the study?
- Were the basic data adequately described?
- Was the statistical significance assessed?
- What do the main findings mean?
- Where are the biases?
- Could there be confounding?
- How are null findings interpreted?
- Are important effects overlooked?
- How do the results compare with previous reports?
- What implications does the study have for you?

The literature review should describe specific gaps in knowledge that the study is intended to fill. The literature review might encompass relevant animal and human experiments, clinical studies, vital statistics, and previous epidemiologic studies. The literature review should also cite the findings of similar studies, and the expected contribution of the current study. Previous findings are useful for the methodological planning of the current study. They may be used to discuss how the findings of the previous research may support the background, significance, research question, hypotheses, and/or design of the proposed study. They may also serve to determine the expected magnitude of the event(s) under study and, if available, in the target population, to characterize the various risk factors for the event and to identify the outcomes and measures that have been used in previous studies. The review assists in providing an assessment of the feasibility of the proposed

study. In addition to seeking information, the review should be a critical appraisal of the evidence in order to assess, analyze and synthesize previous research, and place it in its current context. (Guide on Methodological Standards in Pharmacoepidemiology EMA/95098/2010).

According to Guidelines for Good Pharmacoepidemiology Practices (GPP, 2007) *a description of the research methods* should include:

1. The overall research design, strategy, and reasons for choosing the proposed study design (for example, case-control, cohort, cross sectional, nested case-control, safety trials or hybrid designs)
2. The population or sample to be studied

The population is defined in terms of persons, place, time period, and selection criteria. The rationale for the inclusion and exclusion criteria and their impact on the number of subjects available for analysis should be described. If any sampling from a base population is undertaken, description of the population and details of sampling methods should be provided.

3. The strategies and data sources for determining exposures, health outcomes, and all other variables relevant to the study objectives, such as potential confounding variables and effect measure modifiers

Data sources might include, for example, questionnaires, hospital discharge files, abstracts of primary clinical records, electronic medical records, ad hoc clinical databases, administrative records such as eligibility files, prescription drug files, biological measurements, exposure/work history record reviews, or exposure/disease registries. Use validated instruments and measures whenever such exist, and describe the validation method. If data collection methods or instruments will be tested in a pilot study, plans for the pilot study should be described. Any expert committees and evaluation procedures to be used to validate diagnosis should be described.

4. Clear operational definitions of health outcomes, exposures, and other measured risk factors as well as selection criteria and comparison groups

This section should include clear definitions of: health outcomes, cases, referent groups, exposure and other measured risk factors, selection criteria (exposed and non-exposed persons), description of intervention. It is recommended to use an operational definition that can be implemented independently using the data available in the proposed study. For example „acute hepatitis C"is not an operational definition; a better description would be "hospitalization with a discharge diagnosis of ICD-10 code B17.1"

5. Projected study size, statistical precision, and the basis for their determination

Formal sample size calculations are required for all research studies. Sample size determination depends on precise definition of hypothesis. Describe the relation between the specific aims of the study and the projected study size in relation to each outcome. The power of the study should be considered.

6. Methods used in assembling the study data

Data are the foundation of any empirical study. To avoid any sort of systematic bias in the planning and conduct of an epidemiological study is a fundamental issue, be it information or selection bias. Errors that have been introduced during data collection can in most cases not be corrected later on. Statistical methods are offered to cope with measurement error. An ideal quality of the original data must be the primary goal. Selection bias may be even worse as it cannot be controlled for and may affect both the internal and the external validity of a study. Standardized procedures to ensure the quality of the original data to be collected for a given study are therefore crucial. Pre-testing procedures for research instruments and any manuals and formal training to be provided to interviewers, abstractors, coders or data entry personnel should be described or referenced. The method chosen to collect data depends on the particular exposure to study, the precision of data required, availability of existing records, sensitivity of subject to questioning about the exposure, cost of various methods, etc.

7. Procedures for data management

Data management involves: data collection, entry, verification, data retrieval when required. The aim of data management is to turn information from the subject into a report, efficiently and without errors. Data management is the most important aspect of any epidemiological research. Be sure to describe data management and statistical software programs and hardware to be used in the study. Describe data preparation and analytical procedures as well as the methods for data retrieval and collection.

8. Methods for data analysis

Data analysis includes all the major steps that lead from raw data to a final result, including methods used to correct inconsistencies or errors, impute values, or modify raw data. Data analysis comprises comparisons and methods for analyzing and presenting results, categorizations, and procedures to control sources of bias and their influence on results, e.g., possible impact of biases due to selection bias, misclassification, confounding, and missing data. The statistical procedures to be applied to the data to obtain point estimates and confidence intervals of measures of occurrence or association, for instance, should be presented. Any sensitivity analyses should be described.

9. A description of quality assurance and quality control procedures for all phases of the study

Epidemiologic research on causation uses data in a search for the true nature of the relationship between exposure and disease. Errors that occur during study population selection or in the measurement of study exposures, outcomes, or covariates can lead to a biased estimate of the effect of exposure on risk for the disease of interest. Mechanisms to ensure data quality and integrity should be described, including, for example, abstraction of original documents, extent of source data verification, and validation of endpoints. As appropriate, include certification and/or qualifications of any supporting laboratory or research groups.

10. Limitations of the study design, data sources, and analytic methods

At a minimum, issues relating to confounding, misclassification, selection bias, generalizability, and random error should be considered. The likely success of efforts taken to reduce errors should be discussed.

I. *A description of plans for protecting human subjects*;

This section should include information about whether study subjects will be placed at risk as a result of the study, provisions for maintaining confidentiality of information on study subjects, and potential circumstances and safeguards under which identifiable personal information may be provided to entities outside the study. The need for submitting the protocol to an Institutional Review Board/Independent Ethics Committee and the requirement of informed consent should be considered in accordance with local law.

There is a wide range of documents for protection of human subjects. The Declaration of Helsinki and the provisions on processing of personal data and the protection of privacy as laid down in Directive 95/46/EC and Regulation 45/2001 of the European Parliament and of the Council need to be followed in terms of the ethical conduct of studies. Consideration of ethical issues, data ownership and privacy is an important part of the International Society for Pharmacoepidemiology (ISPE) guideline for Good Pharmacoepidemiology Practices. The main scope of the International Epidemiological Association (IEA) Good Epidemiological Practice (GEP) guideline for proper conduct in epidemiological research is on the ethical principles of field studies, which could also apply to interventional studies, such as the role of ethics committees, patients' informed consent, use and storage of personal data and publication of results. GEP summarize the general ethical principles for research and the important concept of informed consent provides rules for good research behavior under the headings of working with personal data, data documentation, publication, and exercise of judgment with a final note on scientific misconduct. The Uniform Requirements for Manuscripts Submitted to Biomedical Journals by the International Committee of Medical Journal Editors (ICJME) includes clear statements on ethical principles related to publication in biomedical journals addressing authorship and contributorship, editorship, peer review, conflicts of interest, privacy and confidentiality and protection of human subjects and animals in research[9].

J. *A description of plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication;*

There is an ethical obligation to disseminate findings of potential scientific or public health importance. Authorship should follow guidelines established by the International Committee of Medical Journal Editors[10]. The Consolidated Standards of Reporting Trials (CONSORT) statement

9 http://www.ieaweb.org/index.php?option=com_content&view=article&id=15&Itemid=43
10 http://www.icmje.org/.

refers to randomized studies, but provides useful guidance applicable to nonrandomized studies as well[11]. The protocol should specify not only dissemination of results in the scientific media, but also to the community and/or the participants, and consider dissemination to the policy makers where relevant. Publication policy should be clearly discussed, for example who will take the lead in publication and who will be acknowledged in publications, etc.

K. *Resources required conducting the study*;

Describe time, personnel, services (e.g. database access), and equipment required to conduct the study, including a brief description of the role of each of the personnel assigned to the research project.

L. *Bibliographic references;*

List all references, and follow recommended style.

M. *Dated amendments to the protocol.*

Significant deviations from the protocol, such as any changes in the population or sample that were implemented after the beginning of the study, along with the rationale, should be documented in writing. Any changes made after data analysis has begun should be documented as such and the rationale provided.

11 http://www.consort-statement.org/statement/revisedstatement.htm

Chapter 9. General principles of data analysis

Organizing the data
Epidemiologic analysis strategy
    Confidence interval estimation
    Hypothesis testing
    Choosing statistical test
    Two types of error
    Statistical significance or P-value

Organizing the data

The sequence of action in epidemiologic research is similar to any other empirical research area and involves the following major steps: study planning, data collection, data entry, data cleaning, data analysis, interpretation, publication. The data analysis can be divided into **descriptive** and **analytical analysis** (modelling).

According to Rothman K.J. et al. (2008) (Rothman, Greenland & Lash 2008)data analysis has several distinct stages:

- data editing,
- data summarizing,
- estimation,
- interpretation.

In data editing stage, the investigator should review the recorded data for accuracy, consistency, and completeness. In data summarization (data reduction) stage the investigator should summarize the data in a concise form for descriptive analysis, such as contingency tables that classify the observations according to key factors. The summarized data are used to estimate the epidemiologic measures of interest, typically one or more measures of occurrence or effect (such as risk or relative-risk estimates), with appropriate confidence intervals. The estimation stage usually includes statistical hypothesis testing. The interpretation stage is the final step of analysis and involves properly interpreting the results from the summarization and estimation steps (Rothman, Greenland & Lash 2008).

Key words:
descriptive analysis
analytic methods
confidence interval for mean
hypothesis testing
significance testing
power of the study
p value
null hypothesis (h_0)
alternative hypothesis (h_1)
one-tailed test
two sided test
parametric test
non-parametric tests
type I error
type II error

All details of the data analysis should be described in the study protocol. The study should be planned and carried out in such a way that its statistical analysis is able to answer the research questions we are interested in.

The process of data analysis involves four major steps: "cleaning" data, performing analysis, presenting findings and saving your work. Within each of the four major steps, there are four primary tasks: planning your work, organizing your materials, documenting what you do, and executing the work (Long JS 2009).

Planning is important in all types of research. Inadequate planning can lead to misunderstanding about who is doing what. In larger projects it is impossible to remember all the details of what you have done. Planning in principles should cover next topics: general goals (objectives of research), publishing plans, scheduling (timeline with target dates for completing key stages of the project: data collecting, cleaning, documenting data, initial analysis), division of labor (working in a group requires special considerations), datasets (what data, variables will be used, access to restricted datasets, etc.), variable names and labels, data collection and survey design (it is recommended to create a codebook and write command files (for example, do-files in Stata) that create variable and value labels before the start of collecting data), missing data coding, analysis (type of statistical analysis, software), documentation, archiving(Long JS 2009).

Organizing data management involves deciding what happens where, what to name it and how you will find it. The easiest approach to organizing project files is to start with a carefully designed directory structure. Organize folders by subject, not by file type. For a specific project or subproject keep all your main text files, data files and command files in the same folder. Do not mix files from different (sub) projects in the same folder.

If you are collecting data, you need to create a codebook for all the variables. Codebooks describe your dataset. (Long JS 2009). The codebook is the link between the questionnaire and the data entered in the computer, and it should be made early in the process.

The codebook example:

Variable	Source	Meaning	Codes, range	Format
Id	Q 1	Questionnaire number	1-500	3.0
sex	Q 2	Respondent's sex	1 male 0 female 9 no response 1900–2010	1.0
byear	Q 3	Year of birth		4.0
…	…	…		

When analyzing data you will refer to the variables by their names. In some software, variable names can have up to 8 characters; they must start with a letter. Even if your program allows longer variable names, keep them reasonably short. With few variables use names that give intuitive meaning, as in the codebook example. With many variables rather use names derived from question numbers: q3a, q3b etc. In most programs short describing the meaning of the variable can be included in the data set (variable labels) and will be displayed in the output. Always use numerical codes. Not all analyses can handle string codes (e.g. 'M' for male sex), and numerical codes are faster to enter and easier to handle during analysis. Command files that make modifications of your data are vital documentation. When handling information in research you should be able to trace each piece of information back to the original source document ("*audit trail*"). For this reason it is very important to include case identifier (ID) in the original documents and in the data set, all corrections must be documented and explained, all modifications to the dataset (labels added, new variables generated, files merged) must be documented by command files, all analysis must be documented by command files. The purpose is to protect yourself against errors, wasted time, loss of information, to enable external audit and monitoring (Juul S, Frydenberg M 2010).

The research log is the cornerstone of documentation. The log includes dates when work was completed, who did the work, what files were used, and where the materials are located. If you do not document your work, many of the advantages of planning and organization are lost. "It is always faster to document it today than tomorrow" (Long JS 2009).

A log book example (Stata)

Project: Seroepidemiology of XXX			
Working folder: c:\docs\seroxx\data			
Safe folder: c:\docs\seroxx\data\safe			
Input data	Do-file	Output data	Comments
sero1x.rec sero1a.rec	EpiData	sero1a.dta	11.05.2011 Two different persons (X.X., Y.Y) have entered the same data in two separate data files. Two separate data files were compared by EpiData using the function Validate Duplicate Files. Final corrected file sero1a.rec was exported to Stata (sero1a.dta).
sero1a.dta	g_sero1b.do	sero1b.dta	15.05.2011 Added labels to sero1a.dta (X.X.)
sero1b.dta	g_sero1c.dta	sero1c.dta	15.05.2011 Identified errors corrected (see: sero1_correct.doc) (X.X.)
sero1c.dta	g_sero1d.dta	sero1d.dta	16.05.2011 Generated new variables (Y.Y.)

Before entering data you need a complete codebook. All decisions on coding should be made and documented in the codebook before entering data. Examine the questionnaires for obvious inconsistencies before data entry. With small data sets you may enter data in e.g. an Excel spreadsheet or in the SPSS or Stata data window. For larger data sets it is recommended to use specialized data-entry programs, for example EpiData[12]. EpiData enables you to create forms for data entry, to specify valid values and value labels for each variable, and to get a warning during data entry if you enter an outlier (outlying data). EpiData lets you enter data twice (preferable with to different operators) and compare the contents of the two files to get a list of discrepancies. Examine the original source documents, correct the errors in both files and run a new comparison, which should show no errors.

It is a good idea to define variable label and value labels in the codebook before enter data. EpiData can include them in the dataset before the data are entered. In case of receiving the raw data from another source you should define label before examining the results. This command file should be saved and kept in a safe place. After adding labels to data make printouts of a codebook from your data, an overview of your variables and simple frequency tables of appropriate variables. Compare the codebook created with your original codebook and see if you made the label information correctly. Two types of quality-control checks should be performed before beginning the analysis: range checks and logic (or consistency) checks. Inspect the overview table, especially for illegal or improbable minimum and maximum values of variables. Also see if the number of valid observations for each variable is as expected. Inspect the frequency tables (frequencies/ tab1) for strange values and for values that should have labels, but have not. Examine tables that could disclose inconsistencies (Juul S. 2008). If you discover an error, a good practice is to make corrections by command file.

Epidemiologic analysis strategy

After the data have been edited, they are ready to be analyzed. Before starting the analysis you may need to modify data (generate same new variables from primary data, to combine information from several questions, calculate age, etc.). If you modify your data, save the results as a file with a new name. Save command file containing modifications with an appropriate name. It is a good practice to use command files for all analysis. You may use the command window to get experience with single command, but final analysis should be made using command file. Analysis commands and data-management commands should not be mixed in the same command file (Juul S, Frydenberg M 2010).

Before the start of the analysis, data strategy for the analysis should be developed. An analytic strategy that is well planned in advance will expedite the analysis once the data have been edited.

---

12 http://www.epidata.dk

Sequence of an epidemiologic analysis strategy (Gregg 2008):

1. Establish how the data were collected (cohort study, case-control study, etc.) and plan to analyze accordingly.
2. Identify and list the most important variables in light of what you have known about the subject matter, biologically plausible hypothesis and manner in which the study will be (or was) conducted: exposures, outcomes of interest, potential confounders, variables for sub-grouping analysis.
3. To become familiar with the data, plan to perform frequency distribution and descriptive statistics on the variables identified in step 2.
4. To characterize the study population, create tables for descriptive epidemiology.
5. To assess exposure-disease associations, create two-way tables based on study design, prior knowledge and hypotheses.
6. Create additional two-way tables based on interesting findings in the data.
7. Create three-way tables, refinements (e.g. dose-response; sensitivity analysis) and sub-group analysis based on design, prior knowledge and hypotheses, or interesting findings in the data.

After data editing the main steps of an analysis in epidemiology are:

1. Descriptive statistics (descriptive epidemiology: numbers, frequency distribution, percents, rates, etc.).
2. Simple cross-tabulation with appropriate measures of association, tests of significance and confidence intervals.
3. Stratified analysis: evaluate for confounding and effect modification.
4. Multivariate analysis as needed.
5. Interpret appropriately: evaluate for causal relationship.

Descriptive statistics can help to discover data errors and anticipate violations of assumptions required by inferential statistics. It is especially important to know are there any variables that have many missing responses, the range and distribution of each variable and whether there are any outlying values or outliers so that the statistics that are generated can be explained and interpreted correctly. For stratified or sub-group analysis it is important to know if there is sufficient number of responses in each category. Describing the characteristics of the sample also allows other researchers to judge the generalizability of the results.

Continuous data can be summarized in several different ways. One of these are a measures of the center of data distribution (mean, median), and other – measures of the variability of the data (standard deviation (variance), range (minimum, maximum), interquartile range. To summarize continuous data with symmetrical distribution – use arithmetic mean and standard deviation. For continuous data with skewed distribution – use median and interquartile range. For discrete data –

present median. If discrete data have a narrow range, such as stage of disease, it may be better to present the actual frequency distribution to give a fair summary of the data. It is useful to present more than one summary measure for a set of data. If data are going to be analyzed later using methods based on means then it makes sense to present means rather than medians. If the data are skewed they may need to be transformed before analysis and so it is best to present the summary based on the transformed data, such as geometric means (Peacock, Peacock 2011).

In epidemiology, however, the most useful summaries are usually contingency tables in which the frequency of subjects (or units of observation) with specific combinations of variable values is tabulated for the key variables of interest. For continuous variables, the investigator must decide how many categories to make and where the category boundaries should be. The number of categories will usually depend on the amount of data available. If the data are abundant, it is nearly always preferable to divide a variable into many categories. On the other hand, the purpose of data summarization is to present the data concisely and conveniently; creating too many categories will defeat this purpose. For adequate control of confounding, about *five categories* may often suffice. Similarly, if an exposure variable is categorized to examine effect estimates for various levels of exposure, again about five categories may often suffice (Rothman, Greenland & Lash 2008).

Categorical variables are summarized using frequencies in each category together with overall proportions or percentages. Additionally, for ordinal data frequencies can be calculated. Frequency or one-way tables represent the simplest method for summarizing categorical data. They are often used as one of the exploratory procedures to review how different categories of values are distributed in the sample. It is useful to tabulate one categorical variable against another to show the proportion or percentage of the categories of one variable by the other.
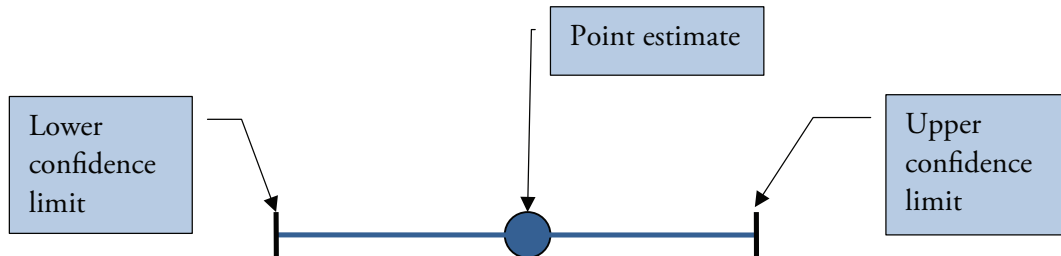
Confidence interval estimation

In research studies it is common to draw conclusions from relatively small amount of data, because it is impractical or impossible to study the whole population. The term statistical parameter refers to an error-free numerical constant that describes a characteristic of a population. We cannot in fact know the value of the parameter exactly, but can estimate it statistically. The calculated statistical estimate will be inexact due to random and systematic error. There will always be an element of uncertainty when we do not have all of the data. Statistical methods based on probability theory are used to quantify this uncertainty. It is common in epidemiology to estimate for example prevalence of condition, e.g. smoking based on the prevalence of condition in a sample. In a given sample the final conclusion may be "35% of the population smokes" (**point estimation**). Estimation takes a form of an interval, such as "between 30% and 40% of population smokes" (**confidence interval estimation**). Interval estimation thus provides more information about the population parameter than the point estimation. We might want to determine whether the prevalence of smoking is different in two groups (**hypothesis testing**). Statistical significance test will help us to weight the

evidence that the sample difference we have observed is in fact a real difference. Note that statistical analysis cannot correct poor study design.

Point estimation is the primary method of data analysis. Interval estimation surrounds the point estimate with a margin of error, thus creating a confidence interval.

Representation of a confidence interval

The width of confidence interval (upper limit minus lower limit) is a measure of estimate's imprecision. Wide confidence intervals indicate low precision and narrow – indicate high precision. Large studies tend to derive narrow confidence intervals (precise estimates). Small studies tend to derive wide confidence intervals (imprecise estimates).

Suppose we selected many samples, then the sample means would follow a distribution known as the sampling distribution of the mean. We could calculate the mean of these sample means and standard deviation. The standard deviation of the sample means is known as the standard error of the mean (denoted by "se", or "SE", or "SEM") and provides an estimate of the precision of the sample mean. Sample mean follows a Normal distribution if the sample size is large. Therefore we can make use of the properties of the Normal distribution when considering the sample mean. In particular, 95% of the distribution of sample means lies within 1.96 standard deviations of the population mean. When we have a single sample, the 95% **confidence interval** (CI) **for the mean** is:

Sample mean – (1.96 x SEM) to Sample mean + (1.96 x SEM), were $SEM = \dfrac{SD}{\sqrt{n}}$

95% is the most commonly used percentage for CIs and the multiplier is 1.96 for large samples. For a sample mean, a sample size of 100 is considered large. 95% CI is a range of values which has a 95% probability of containing the true population value (parameter) in the sense that if an infinite number of samples were drawn to estimate the value of interest, 95% of their 95% CIs would contain the true population value.

Other percentages can be used such as 90% or 99%. 90% CI has a probability of 90% of containing the true value and uses the multiplier 1.64. 99% CI has probability of 99% of containing the true value and uses the multiplier 2.58.

The sampling distribution of a **proportion** follows a Binomial distribution. If the sample size, *n*, is reasonably large, then the sampling distribution of the proportion is approximately *normal*. Taking into account *normal* distribution we can calculate the standard error of a proportion and then estimate the 95% confidence interval for a sample proportion. Population proportion we estimate by $p = r/n$ (where *r* is the number of individuals in the sample with the characteristic of interest), and its standard error is estimated by:

$$SE(p) = \sqrt{\frac{p(1-p)}{n}}$$

The 95% confidence interval for the proportion from large sample is estimated by:

$$p - 1.96 \times SE(p) \text{ to } p + 1.96 \times SE(p)$$

If the sample size is small (less than 5) then we have to use the Binomial distribution to calculate exact confidence intervals.

A 95% confidence interval for proportion is a range of values which has 95% probability of containing the true population proportion. In other words, we have 95% confidence that the true value of the proportion in the population from the sample was taken lies within the interval (Peacock, Peacock 2011).

Hypothesis testing

We often gather sample data in order to assess how much evidence there is against a specific hypothesis about the population. We use a process known as **hypothesis testing** (or **significance testing**) to quantify our belief against a particular hypothesis. There are always two mutually exclusive hypotheses since, if the hypothesis being tested is not true, then the opposite hypotheses must be true. A measure of the evidence for or against the hypothesis is provided by ***P* value**.

We usually test the **null hypothesis** (H_0) which assumes *no effect* (e.g. the difference in prevalence equals zero) in the *population*. For example, if we are interested in comparing smoking prevalence in men and women in the population, the null hypothesis would be H_0: smoking prevalence are the same in men and women in the population. We then define the **alternative hypothesis** (H_a). The alternative hypothesis relates more directly to the theory we wish to investigate. So, in the example, we might have H_a: the smoking prevalence is different in men and women in the population. In an example, the alternative hypothesis is general and allows the difference to be in either direction. This is known as a two-sided or two-tailed test. In some, very rare, circumstances, we may carry out a **one-tailed test** in which a direction of effect is specified in H_a. This might apply if we are considering, for example, a disease from which all untreated individuals die (a new drug cannot make things worse). One-sided test does not distinguish between "no difference" and a

"harmful effect" of the new treatment. **Two-sided tests** should always be used unless there is clear justification at the outset to use a one-sided test (Peacock, Peacock 2011).

Steps in doing a significance test (adapted from Bland M.) (Bland 1987):

1. Specify the hypothesis of interest as a null and alternative hypothesis
2. Decide what statistical test is appropriate
3. Use the test to calculate the P value
4. Weight the evidence from the P value in favor of the null or alternative hypothesis.

Both the null hypothesis and the alternative hypothesis should be specified in advance. When little is known about the association being tested, you should specify a null hypothesis that the exposure is not related to disease (e.g., RR = 1.0 or OR = 1.0). The corresponding alternative hypothesis states that exposure and disease are associated (e.g., RR≠1.0 or OR ≠ 1.0). Note that this alternative hypothesis includes the possibilities that exposure may either increase or decrease the risk of disease. (Gregg 2008).

Deciding which statistical test to use depends on the design of the study, the type of variable and the distribution that the data being studied follow.

Choosing statistical test

Hypothesis tests which are based on knowledge of the probability distributions that the data follow are known as **parametric tests.** Often data do not conform to the assumptions that underlay these methods. In these instances we can use **non-parametric tests** (sometimes referred to as *distribution-free* tests, or **rank methods**). These tests generally replace the data with their ranks (i.e. the numbers 1, 2, 3 etc., describing their position in the ordered data set) and make no assumptions about the probability distribution that the data follow. Non-parametric tests are particularly useful when the sample size is small (so that it is impossible to assess the distribution of the data), and/or when the data are measured on a categorical scale. However, non-parametric tests have less power to detect a real effect than the equivalent parametric test if all the assumptions underlying the parametric test are satisfied. (Petrie, Sabin 2005).

Choosing a statistic when there is one outcome variable only (pasted from Peat J, Barton B 2005).

Type of variable	Number of times measured in each participant	Statistic	SPSS menu
Binary	Once	Incidence or prevalence and 95% confidence interval (95% CI)	Descriptive statistics; Frequencies
	Twice	McNemar's chi-square Kappa	Descriptive statistics; Crosstabs

Continuous	Once	Tests for normality	Non-parametric tests; 1 sample K-S Descriptive statistics; Explore
		One sample t-test	Compare means; One-sample t-test
		Mean, standard deviation (SD) and 95% CI	Descriptive statistics; Explore
		Median and inter-quartile (IQ) range	Descriptive statistics; Explore
	Twice	Paired t-test	Compare means; Paired-samples t-test
		Mean difference and 95% CI	Compare means; Paired-samples t-test
		Measurement error	Compare means; Paired-samples t-test
		Mean-versus-differences plot	Graphs; Scatter
		Intraclass correlation coefficient	Scale; Reliability Analysis
	Three or more	Repeated measures ANOVA	General linear model; Repeated measures

Choosing a statistic when there is one outcome variable and one explanatory variable (pasted from Peat J., Barton B. 2005) (Peat J 2005).

Type of outcome variable	Type of explanatory variable	Number of levels of the categorical variable	Statistic	SPSS menu
Categorical	Categorical	Both variables are binary	Chi-square Odds ratio or relative risk Logistic regression Sensitivity and specificity Likelihood ratio	Descriptive statistics; Crosstabs Descriptive statistics; Crosstabs Regression; Binary logistic Descriptive statistics; Crosstabs Descriptive statistics; Crosstabs
Categorical	Categorical	At least one of the variables has more than two levels	Chi-square Chi-square trend Kendall's correlation	Descriptive statistics; Crosstabs Descriptive statistics; Crosstabs Correlate; Bivariate
Categorical	Continuous	Categorical variable is binary	ROC curve Survival analyses	Graphs; ROC curve Survival; Kaplan-Meier
Categorical	Continuous	Categorical variable is multi-level and ordered	Spearman's correlation coefficient	Correlate; Bivariate
Continuous	Categorical	Explanatory variable is binary	Independent samples t-test Mean difference and 95% CI	Compare means; Independent-samples t-test Compare means; Independent-samples t-test
Continuous	Categorical	Explanatory variable has three or more categories	Analysis of variance	Compare means; One-way ANOVA
Continuous	Continuous	No categorical variables	Regression Pearson's correlation	Regression; Linear Correlate; Bivariate

Parametric and non-parametric equivalents (pasted from Peat J., Barton B. 2005)(Peat J 2005).

Parametric test	Non-parametric equivalent	SPSS menu
Mean and standard deviation	Median and inter-quartile range	Descriptive statistics; Explore
Pearson's correlation coefficient	Spearman's or Kendall's correlation coefficient	Correlate; Bivariate
One sample sign test	Sign test	SPSS does not provide this option but a sign test can be obtained by computing a new constant variable equal to the test value (e.g. 0 or 100) and using non-parametric test; 2 related samples with the outcome and computed variable as the pair
Two sample t-test	Wilcoxon rank sum test	Non-parametric tests; 2 related samples
Independent t-test	Mann-Whitney U or Wilcoxon Rank Sum test	Non-parametric tests; 2 independent samples
Analysis of variance	Mann-Whitney U test	Non-parametric tests; K independent samples
Repeated measures analysis of variance	Friedmans ANOVA test	Nonparametric tests; K independent samples

Two types of error

Since significance tests use sample data to make inference about populations, using the results from a sample may lead to wrong conclusion when we reject/do not reject the null hypothesis. There are two types of error. Type I error: we reject the null hypothesis when it is true, and conclude that there is an effect when, in reality, there is none. The maximum chance (probability) of making a Type I error is denoted by α (alpha). We usually set a limit of 0.05 (5%) for the probability of a type 1 error, which is equivalent to 0.05 cut-off for statistical significance. Type II error: we do not reject the null hypothesis when it is false, and conclude that there is no effect when one really exists. The chance of making a type II error is denoted by β (beta); It is widely accepted that the probability of type II error should be no more than 0.20 (20%); $1 - β$ is the power of the test. The power, therefore, is the probability of rejecting the null hypothesis when it is false; i.e. it is the chance (usually expressed as a percentage) of detecting, as statistically significant, a real effect of a given size.

Statistical significance or P-value

All test statistics follow known theoretical probability distributions. We relate the value of the test statistic obtained from the sample to the known distribution to obtain the P-value, the area

in both (or occasionally one) tails of the probability distribution. Most computer packages provide the two-tailed *P*-value automatically. The *P*-value is the probability of obtaining our results, as extreme or more extreme than observed, if the null hypothesis is true. Using conditional notation, $P = Pr(data|H_0)$. The null hypothesis relates to the population of interest, rather than the sample. Therefore, the null hypothesis is either true or false and we *cannot* interpret the *P*-value as the probability that the null hypothesis is true. 0.05 (5%) is commonly used as a cut-off. P<0.05 is commonly described as statistically significant and P≥0.05 is described as not statistically significant. It is best always to report the exact P value from a test than report findings as P<0.05 rather than report findings as P≥0.05 or worse "P=NS"(meaning non-significant). Not significant does not mean: "there is no difference" or "there is no effect". It means that there is insufficient evidence for a difference or effect. Non-significance may reflect no association in the source population but may also reflect a study size too small to detect a true association in the source population. Statistical significance does not by itself indicate a cause-effect relationship. An observed association may indeed represent a causal relationship, but it may also be due to chance, selection bias, information bias, confounding, and other sources of error in the design, execution, and analysis of the study. Statistical testing relates only to the role of chance in explaining an observed association, and statistical significance indicates only that chance is an unlikely (though not impossible) explanation of the association (Gregg 2008).

Confidence intervals and hypothesis tests are closely linked. The primary aim of a hypothesis test is to make a decision and provide an exact *P*-value. A confidence interval quantifies the effect of interest (e.g. the difference in means). The *p*-value does not tell us about the direction of association, magnitude of association, precision around the point estimate. Many epidemiologists prefer confidence intervals to significance tests when conducting statistical inferences. Statistical tests they used in a confirmatory manner.

Chapter 10. Writing a manuscript

Guidelines for manuscript preparing
Tabular & graphical data presentation

Guidelines for manuscript preparing

Writing a research paper should be a straight forward exercise that translates scientific data into a clear, practical lesson for the specialist or community depending on a research question. Moreover, it demands a constant training and discipline to learn how to convey a scientific message in a logical manner understandable to the audience that is targeted to. Before starting writing it is always useful to remember the practical things postulated by Dodson TB in 2007 (Dodson 2007):

- There is no such thing as a paper that is too short
- Write short declarative sentences, they contain more meaning than long ones
- All studies can be simplified into 2x2 table
- The purpose of writing is to communicate what you have learned to the readers
- It is easier to write when you have something to say
- A good article is one that you would like to read yourself

Key words:
title
abstract
introduction
methods
results
discussion
one / two way frequency
tables
histograms
bar charts
box plot
scatter plot
line graphs

The reporting of biomedical research, be it public health or laboratory investigation research, falls under certain rules that are most often presented for the potential author as the instructions for the authors on the website of a certain journal. Those rules are agreed between the main journal publishers and are very much alike with few nuances being different. The problem evolves with judging what issues in particular should be covered in each section of an article. For being confident in proper reporting your research findings, we strongly recommend to appeal for STROBE statement presenting the guidelines for reporting the observational studies (von Elm et al. 2008). Since a lot of medical research is performed in observational studies and reporting for results is inadequate or even mistaken, this group seeks for accurate and complete report. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement is freely available on the website

of Journal of Clinical Epidemiology[13] and we strongly recommend to follow them. The checklist of items that should be addressed in reports of observational studies are as follows:

I.   Title and abstract

II. Introduction covering the background and objectives of the study

III. Methods: study design, setting, participants, variables, data sources/measurement, study size, statistical methods

IV. Results: participants, descriptive data, outcome data, main results, other analysis

V.   Discussion: key results, limitations,bias, interpretation, generalizability

To get the better perception of how it looks in particular case we suggest following this cross sectional/observational article with annotations echoing the main recommendations of STROBE statement (Dadoniene et al. 2007).

Title may contain the main findings of the study

Rheumatoid arthritis in Lithuania: Need for external help from the onset of disease

Abstract

Abstract is an informative and balanced summary of what was done and what was found.

Purpose. To estimate the burden of rheumatoid arthritis (RA) in Vilnius, Lithuania, the former socialist country in Eastern Europe, in terms of patients' need for help from other persons and to explore the factors which influence the need for physical help. *Method*. Some 537 patients with RA, registered in Vilnius, answered questions about socio-demographics, disease characteristics, categories of required help, the use of major appliances and adaptations, underwent a clinical examination and filled in the modified health assessment questionnaire (MHAQ) and arthritis impact measurement scale (AIMS). Logistic regression was used to assess which variables from those explored influenced the need for physical help. *Results*. A total of 230 (42.9%) patients out of 537 were requiring help from other persons, and the proportion was equally high in all the disease duration categories. A quarter of the patients (25.1%) were classified to ACR III and IV functional impairment groups. In multivariate logistic regression model the risk to become dependent on external help ultimately depended on MHAQ (10.32 [CI 95% 6.57; 16.23], $p < 0.001$) but the use of joint stabilization measures (1.97 [CI 95% 1.06; 3.64], $p < 0.01$) and 28 tender joints count (1.02 [CI 95% 1.0; 1.06], $p < 0.05$) were also important. *Conclusions*. Nearly half of the patients reported being dependent on others and a quarter of patients were in definite need for that. The functional impairment is the most important risk factor, although identifying the group using joint stabilization measures routinely may be of practical value in order to define the risk group which may need the external help in future.

Keywords: Rheumatoid arthritis; functional impairment; disease burden

Introduction

Preservation of physical function is one of the most important issues in the long-term outcome of rheumatoid arthritis (RA) **1** which greatly determines the possibi-
[Explain scientific background]
ding an independent life. It is acknowledged
atients' community but still under-looked by
essionals. Though the burden of the disease is
be analysed from different points of disability, the most important data, from the patients' viewpoint, relating to the needs for external help, the range of the needs and the disease-related factors possibly influencing them, are underestimated **2**. Moreover, the disease burden may be related to the economical background
[You may need to explain the rationale of the study]
untry since the social and health care systems
tly from country to country. Lithuania is one
ormer socialist countries in Eastern Europe,
s started the process of economic transition

Health care and rehabilitation system in Lithuania

After Lithuania regained its independence, the medical community started looking for effective ways to improve the health care system and quality of provided services. In 1991, the development of primary health care, the introduction of the family doctors' institution and categorizing the health care-providing institutions according to the primary, secondary and tertiary medical service provision level was set. The Health Insurance Law was approved in 1996 in Lithuania. It foresees a compulsory health insurance for all permanent residents in Lithuania and is executed by one state institution – the State Patients' Fund (SPF). An additional private insurance is foreseen in the Health Insurance Law, however it is still not popular and used by a small part of the population with high income. Those who are not insured may apply only for necessary medical aid; other services should be paid under the prices set by Ministry of Health. For those insured, the SPF fund covers visits to the doctor, treatment at the hospital and partial medical rehabilitation at policlinics. The partial

from a state-planned to a market-oriented economy in 1991, but frequent changes in governments made the economical reforms inconsistent and tardy, resulting in relatively lower GDP and higher unemployment if compared to western European countries. Musculoskeletal disorders, and RA among them, are the main and the major cause of permanent functional disability in disability structure forming a part of 19.2% in year 2004 in Lithuania. It may be thought that this segment of population may need broader range of health and social services though the satisfaction of the needs is not explored yet. The objective of this study was to estimate the RA burden in Lithuania (Vilnius) in terms of patients' need for external help and to elucidate
[State clear the objective]
influencing the need for physical help.

reimbursement for some medicines according to the categorized list of diseases and medicines and orthotic devices for patients with special needs and functional impairment are also provided by patients' funds. For example, an RA patient can apply for 50% reimbursement for disease modifying and one non-steroidal drug at a time. The reimbursement for musculoskeletal medicines constitutes a fraction of 2.66% among all SPF expenditure for medicines. The endoprosthetic operations are fully covered by SPF, although the waiting time takes 2 – 5 years. The waiting time may be shortened by the patient paying him/herself for the artificial joint. The rehabilitation at health resorts for whatever reasons is not covered by SPF, although some physiotherapeutic procedures are available at policlinics and are often prescribed by rehabilitation specialists. In 2004, SPF spent 158.8 Euro on average per insured person. The GDP of Lithuania in year 2004 was 23,702 million US dollars (app. 2,000 US dollars per capita) and the fraction used by health sector made about 5% at that time.

Material and methods

> Give the eligibility criteria, the sources and methods of selection of participants

> Present key elements of study design early in the paper

> Explain how study sample was arrived at

Criteria for the enrolment comprised [...] RA according to the ACR 1987 revised [...] and a residential address in Vilnius. T[...] established in 1998 for research purpo[...] survey included 1,018 patients with RA, and w[...] scribed [...] elsewhere 3. The data here pre[...] ed were [...] ined from the cross sectional analysis [...] rried out [...] November 2000 to November 2004 Some 923 patients were asked by mail to participate in the study and 620 agreed to and finally 537 were interviewed and examined, accounting for a response rate of 58.2%. The interview comprised socio-demographic questions,

> Clearly define all the variables: outcomes and exposures; and details of the method of asssessment

[...]ce of co-morbidities (pulmonary, renal, thyroid, [...] and vessel diseases, peptic ulcers, diabetes, [...]res, low back pain, stroke, psychiatric diseases and [...]), and extra-articular RA manifestations, use of [...]RD(s), endoprosthetic operations and patients' [...] conditions. During the clinical examination, [...] strength and 6 m walking time were measured; 28 tender and 28 swollen joints were counted. Patients were particularly asked about the daily use of joint stabilization measures (orthotic devices, splints, taping) and about the use of major appliances and adaptation

measures from the list. They were also asked to fill in a visual analogue scale (VAS) for pain, arthritis impact measurement scale (AIMS) and modified health assessment questionnaire (MHAQ). Finally, they were asked about the need for physical help from others as a separate question and, if present, to structure them according to the eight domains from the MHAQ. To compare the characteristics among the three distinguished disease duration categories, chi-square was used for proportions and ANOVA for continuous variables. Self-reported help-dependent versus not help-dependent group were compared on predictor variables. Spearman correlations between predictor variables were examined in order to gauge their potential multicollinearity and usefulness in future regressions. Only the variable[...] statistically significant in the univariate analysis an[...] correlating less than 0.5 were entered into the mo[...]. The variables being not significantly di[...] the two groups or correlation index s[...] an 0.5 were not included into furt[...] ltivariate logistic regression was applied to assess the risk for external help needed. The level of significance was set at 0.05.

> Describe all statistical methods and set the p value.

Results

Table I shows the socio-demographic and disease characteristics of 537 RA patients in different disease duration categories. The pain VAS and physical functioning measurements, including 6 m walking time, grip strength and MHAQ, significantly worsened over years while disease characteristics – swollen, tender joint count, the proportion of extra-articular manifestations and co-morbidities – remained stable over time. In total, 37.4% reported MHAQ score higher than 1.5. It is of note that only half of the patients were taking DMARD(s) at interview and the proportion of endoprosthetic operations performed was strikingly low. One third of RA patients (31.2%) used major appliances and adaptations in their daily activities among which cane was used in 109 cases among 167 cases and joint stabilization measures were used by a

still lower proportion of the patients (17.0%). Some 47 patients out of 91 were helping knees, followed by splints used for wrists and hands [...] 19 cases, and [...] parts of the body being supported less fre[...] assistive devices and joint stabilization measures [...] increasingly found in longer disease duration g[...] A total of 230 patients (42.8%) reported the ne[...] external help, and the proportion was equally high in all disease duration categories. Of those, 41 (17.8%) people were living alone. We found 3.9% of patients bedridden or in a wheelchair. When comparing the groups dependent and not depend[...] on external help in univariate statistics, the majority of so[...]den[...] and disease variables differed between [...] except for age, gender, disease duration, pr[...] of endoprosthesis operations and DMARD[...]

> Report numbers of outcome events or summary measures

> Report any other analyses done: analyses on subgroups and their interactions.

which were excluded from further analysis (Table II). After examining for bivariate correlations, five more variables were excluded: Pain VAS, grip strength, 6 m walking time, the use of major appliances and adaptations because of interactions with MHAQ, and AIMS score for anxiety strongly correlating with AIMS for depression. Therefore, a ten variable logistic regression model was built. The possibility of becoming dependent on external help is ultimately dependent on MHAQ (10.32 [CI 95% 6.57; 16.23], p < 0.001), although other factors may predict the need for external help in future to some extent, particularly the use of joint stabilization measures (1.97 [CI 95% 1.06; 3.64], p < 0.01) and 28 tender joints count (1.02 [CI 95% 1.0; 1.06], p < 0.05) (Table III). Identifying the group using orthotic devices to support joints routinely may be of practical value in order to define the risk group which might need the external help in future.

Discussion

In the first paragraph summarize key reults with reference to study objectives.

Our study, based on Vilnius RA register, elucidated the proportion of patients who experience the most severe outcomes of disease being unable to lead an independent life. Nearly half of the patients need help from others and a quarter of all of the patients are in definite need of that. This is the first study from an Eastern European country reporting on RA burden based on a representative sample and a clearly defined population.

Give a cautious overall interpretation of results. Compare your results to other studies KEEP IT FOCUSED !

Relatively little is known about the burden of RA in terms of the proportion of patients requiring help, to what extent they need help, how the help-dependent patients manage in their everyday lives, and what are the patients' future help perspectives, although it goes along with the most important need from the patients' position – to stay independent in one's life 4. To the best of our knowledge, quite a few extensive studies on the loss of independence were published by Westhoff et al. based on Berlin rheumatological database 5. Generalizing the results to all RA patients treated by rheumatologists in Berlin, 33% were expected to be dependent on external help and 7% on care. Our study elucidated the proportion of 42.8% for patients with a self-reported need for external help and a quarter of patients in ACR III and IV class (25.1%). The proportion of patients with the need for definite or self-reported help from other European studies could not be sought directly. Based on the prediction made by same authors, that half of the patients with HAQ may require definite help, it could be assumed that may be around 20% in our study and almost the number of patients in Oslo 6 but a less proporti... the Finnish community 7. This early study found...

one of 103 patients bedridden (ARC IV). The total joint replacement surgery performed for 19% of patients was given as an explanation of the absence of marked restriction in mobility. In agreement with this, the recent Finish study reported a low median HAQ score (0.7) of a large cohort of RA patients predicting of a very small proportion of patients in severe functional impairment. Assuming that the patients in ACR class III and IV are dependent on other people's help, Italian 8 and French 9 studies, both hospital-based studies, indicated nearly half of the patients in definite need for help, 49% and 53%, respectively, contradicting the above-mentioned studies. Our study revealed a considerably lower proportion (25.1%) of all the patients in ACR class III and IV. Therefore, our study showed similar results in definite need for external help with Northern European countries but less comparable with South European countries. The explanatory factors mostly lie in the different methodological approach of the study performed. The results from community studies like Berlin, Oslo, and this study may differ greatly from hospital centre-based Italian and French studies with a higher proportion of functional impairment within the latter studies, though the different genetic backgrounds of different settings of patients should not be ruled out. The measures taken towards the reduction of the

Discuss limitations of the study, taking in account sources of potential biases or imprecision.

... burden encounter the assistive measures used at ... and individually, and their analysis is worthwhile ...ming that findings are solitary 10. The assistive ...es are generally perceived as having a positive ... on function since they do not interfere with work ...rmance, endurance and difficulty but the effect on

pain level is limited 11. According to van der Heidje et al., the use of devices reflect a distinct aspect of physical function than self-reported functional disability itself, and deserve a distinct approach 12. The proportion of patients using them differs depending on the study place. The study by Veehof et al. showed the majority of patients (78%) possessing at least one or more assistive devices in The Netherlands and Germany 13, while our study managed to reveal 31.2% using assistive devices and only 17% helping joints. The most likely explanation and in accordance with the quoted study are differences in health care systems for prescription and reimbursement of assistive devices with a small proportion being reimbursed for persons completely depending on care. This study showed that use of joint stabilizing measures did not prevent RA patients from staying independent from external help as it may be thought. On the contrary, it was shown that RA patients using joint stabilization measures are more likely to depend on others and this may serve as a predictor for possible help dependence, unlikely to the use of

technical aids being not related to external help needed. The information on physical functioning and joint stabilization measures taken together may be important to elucidate the group of those needing exclusive social support. The results from this study may reflect the current situation in the entire Eastern Europe because of similarities in health care systems, and can be used for comparisons of the RA burden between different countries with different access to health care facilities. Nevertheless, a methodologically consistent approach to the cross-country comparisons on RA burden is needed 14. In conclusion, nearly half of the patients reported as being dependent on others, and a quarter of patients are in definite need of that. The functional impairment is the most important risk factor for that, although identifying the group using join stabilization measures routinely may be of practical value in order to define the risk group which

References

References are quoted as agreed by International Committee of Medical Journal Editors for uniform requirements for manuscripts submitted to Biomedical Journals. Don't do it by hand, use an appropriate reference data software.

Table I. **The demographic, disease, need for help characteristics of 537 Lithuanian rheumatoid arthritis patients (mean [SD] for continuous variables, *n* (%) for dichotomous variables**

	Disease duration < = 5 (*n* = 199)	Disease duration 6 – 10 (*n* = 123)	Disease duration >10 (*n* = 215)	Total (*n* = 537)
Female	166 (83.4%)	109 (88.6%)	186 (86.5%)	461 (85.1%)
Education in years	12.2 (4.1)	12.1 (4.8)	11.8 (4.6)	12.0 (4.5)
Pain VAS (1 – 10)	5.4 (2.3)	5.8 (2.3)	6.3 (2.8)	5.9 (2.3)**
28 – Swollen joints	6.1 (6.9)	5.7 (6.4)	5.6 (6.6)	5.8 (6.7)
28 –Tender joints	13.4 (9.5)	13.1 (9.7)	13.3 (12.2)	13.3 (10.7)
Walking time 6 m, sec	7.6 (4.4)	9.3 (5.6)	10.0 (6.1)	9.0 (5.5)***
Grip strength (right hand), mmHg	86.4 (38.7)	80.7 (39.3)	73.7 (37.4)	80.0 (38.7)**
MHAQ	1.2 (0.8)	1.3 (0.7)	1.5 (0.8)	1.3 (0.8)
Endoprosthesis operations	2 (1.0%)	6 (4.9%)	20 (9.3%)	28 (5.2%)
Use of major appliances and adaptations	38 (19.1%)	38 (30.9%)	91 (42.5%)	167 (31.2%)***
Use of joint stabilization measures	21 (10.6%)	25 (20.3%)	45 (21.0%)	91 (17.0%)*
Need to climb stairs every day	105 (53.3%)	57 (47.1%)	93 (43.9%)	255 (48.1%)
Living alone	37 (18.6%)	21 (17.1%)	46 (21.4%)	104 (19.4%)
Need for external help	**76 (38.2%)**	**54 (43.9%)**	**100 (46.5%)**	**230 (42.8%)**

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Tabular and graphical data presentation

Tabular and graphical data presentation helps the reader to perceive them in easiest and time consuming way. Moreover, the easiest way to write the result section is to complete all of the tables and graphs first. A manuscript usually contains up to six tables and a few graphs.

There are several most popular options to present the data in tables and graphs. The *one* and *two way frequency* tables and several types of figures, in particular *bar charts, histograms, box plot, scatterplot and line graphs* are presented in this section based on the data presented earlier in the quoted article.

A **one way** frequency table shows the result of the tabulation of observation at each level of a variable. It means we show the variable(s) not linked to others. Usually they are so called the descriptive tables that are suitable for presenting the basic characteristic of the study group but do not show any relation between them. The variables used in frequency tables may be nominal, ordinal, or continuous. When continuous variables are used in tables, their values are often grouped into categories or less often they are quoted as mean and standard deviation. Computer packages can be used to categorize continuous variables (recoding) and to tabulate the data into one or two way tables. Discrete data (nominal or ordinal) with few categories are presented in proportion and absolute data, of course.

Two way frequency tables, formed by the cross tabulation of two variables, where the rows represent the categories of one variable and the columns represent the categories of a second variable. They are usually more interesting than one way tables because they show the relation between the variables and contain the descriptive information equally. It is always easier to construct the two way table when the grouping data are nominal or ordinal, but if not the most useful approach with continuous data is to group them first, and then to construct a frequency distribution of the grouped data. For example, disease duration is often categorized into year intervals and cross tabulated with other variables under the study, gender like it is in table 10.1.

Table 10.1. The cross tabulation of grouped disease duration and gender of 537 rheumatoid arthritis patients showed in a two way frequency table

| Disease duration in years | Gender | | Total (%) 537 |
	Female N=461 Counts, (%)	Male N=76 Counts, (%)	
Up to 5 years	166 (36)	33 (43.4)	199 (37.1)
6–10 years	110 (23.9)	14 (18.4)	124 (23.1)
> 10 years	185 (40.1)	29 (38.1)	214 (39.9)

Table I of the quoted article is also a two way table since it cross tabulates disease duration with other particular variable and shows relation between the disease duration and one particular variable.

It is also worth of noting that the title of the table should contain sufficient information to allow the reader to understand the table without referring to the text

in the article. The total sample size, the group size, and the origin of the data and the dimensions used should be clearly stated within the table. If the table aims to compare the data between the groups, the p value level should be stated in a separate column or under the same table.

The graphical presentation of the data is useful when there is a need to emphasize some important information that should not be missed by reviewers and readers. Sometime it may double the information given in tabular format. We all know what a pie chart is, it is simple to sketch and understand it because each segment of a pie chart is proportional to the frequency of the category it represents. For better understanding percentage values should be always displayed together with counts. A disadvantage of a pie chart is that it can only represent one variable and we will need a separate pie chart for another variable. Moreover it can lose clarity if it is used to represent more than five categories. The **simple bar charts** are an alternative to pie chart for nominal data. This is a chart with frequency or proportion on the vertical axis and category on the horizontal axis and shows the number or proportion of people by levels of a nominal or ordinal variable (Figure 10.1). In bar charts, the length of the bar shows the number of observations or the value of the variable of interest for each level of the nominal or ordinal variable. The widths of the bar are the same for all the levels of the nominal or ordinal variable, and the width has no meaning! The **clustered bar chart** of gender distribution in different disease duration categories from the Table I is presented in the Figure 10.2 and reflects data exactly as it is presented in the table. The length of the bar shows the number of female and men, while the widths represent the separate category of disease duration.

Bar charts can also be used to present more complicated data. The tabulated data in two way tables can be presented in a bar chart format. For instance, the most important data about need for help (the status of *need* is yes or no) in different disease duration groups may be presented by three **stacked bars** each of which is presenting 100 per cent group divided by the need of help. The increasing need for help in each category can be observed. Three stacked bars represent data from the last line in the Table I exactly as they are presented in the columns (Figure 10.3).

The graphical presentation of continuous data is more complicated than nominal or categorical. A continuous variable can take a very large number of values, so it is usually impractical to plot them without first grouping the values. The grouped data is plotted using a **histogram**. It is a graphic representation of the frequency distribution of a variable. A histogram is similar to a bar chart but is used for continuous data. The values are grouped into intervals (often called bins or classes) that are usually of equal width. Rectangles are drawn above each interval, and the area of rectangle represents the number of observations in that interval. The height of the interval, as well as its area, represents the frequency of the interval. In contrast to bar charts, there are no spaces between the rectangles unless there are no observations in some interval. We demonstrate here the construction of a histogram for the data of age of 537 (Figure10.4). How many intervals should there be and how large should the intervals be? There are no prompt answers to these questions and sometimes mathematical calculations are needed but the general guidelines are presented here. Generally 5 to 15 intervals would be used, with a smaller number of intervals used for smaller sample size. The width of the interval is also arbitrary; in this case we chose the interval of 10 years. Equal

size intervals are used in most histograms. Depending on the number of intervals and how their boundaries are set, it is possible for histograms constructed from the same data to have different shapes. However, histograms say basically the same things about the distribution of the sample data even though their shapes are different. In addition, most of computer based programmes add normal distribution curve to the histograms making it easier to judge if the data fit the normal distribution.

Another way of presenting continuous data is **the cumulative frequency curve.** It gives us an opportunity to estimate the cumulative frequency for any value on the horizontal axis (Figure 10.5). By drawing a line vertically upwards from value 60 years upwards to the curve, and then horizontally to the vertical axis, one can see that about 50 % of the patients are younger than 60 years.

The box and whisker plot, or just **box plot**, graphically gives the approximate location of the median, quartiles and extreme values. The advantage of using box plots when exploring data is that several of the characteristics of the data such as symmetry features, the range, and dispersion of the data can be easily compared between the different groups. The lower/left and upper/right ends (hinges) of the box mark the 25th and 75th percentiles or the location of the first and third quartiles, while the solid band indicates the 50th percentile or the median. The whiskers represent the range of values. If the box plot is presented vertically, the area from the top edge to the bottom edge of the box represents the interquartile range. From the summary statistics of the modified health assessment questionnaire data (MHAQ) we found the following information:

- Lowest value = 0.00
- First quartile = 0.75
- Median = 1.25
- Third quartile = 1.86
- Highest value = 3.00

These values are plotted in a box plot in Figure 10.6. We can use this figure to assess the symmetry of the health assessment data. The box plots and histograms give us an indication of whether or not the data are skewed. For these patients the distance from the median to the third quartile and further to the highest value looks longer than to the lowest value, indicating the distribution is slightly skewed to the right. Similarly the box plots can be constructed and compared between two and more groups but within one variable. The box plot in Figure 10.7 fully represents data from the Table I in the manuscript showing the health evaluation in different disease duration categories and reflecting how it increases (meaning becoming worth) in parallel with disease duration becoming longer.

The two-dimensional **scatter plots** are analogous to the two way frequency table in that it facilitates the examination of the relation between two variables. Unlike the two-way table, the two-dimensional scatter plot is most effectively used when the variables are continuous. The scatter plot pictorially represents the relation between two continuous variables. In a scatter plot, a plotted point represents the value of two variables for an individual. In Figure 10.8 we examine the relationship between MHAQ and disease duration of 537 patients using the scatter plot. Overall, the scatter

plot suggests weak if any relationship between variables. There is a positive association between the variables when larger (smaller) values on one variable appear with larger (smaller) values of the other variable. The association would be negative if the individuals with large values of one variable tended to have small values of the other variable and conversely. Scatter plots are most effective for small to moderate sample size. When the samples are large as it is in our case, the circles of the data are overlapping and information about each of the point may be disappearing.

When there are many variables and the relationship between each of the pairs is important the scatter plot matrix can be useful in displaying multiple two-way scatter plots.

Tips about regression

The scatter plot that is usually built up to emphasize the relationship between the two variables, most often is constructed together with the regression line as it appears in Figure. 10.8 and is called *simple linear regression*. Actually it is an extension of correlation coefficient and it can be calculated from linear regression (Bowers 2008). There are two very important issues when talking about simple linear regression. The relationship may not always be linear, though the computer based program can build a line on the scatter plot easily. It may be exponential, asymptotic, cyclic curve etc. and application of linear equation for describing the relationship between the two variables would be mistaken. To judge whether linear relationship is applicable, what we would like to in majority of situations (be it laboratory clinical or public health data), the diagnostics of residuals should be applied. And what are residuals? The distance of each point in the scatter from the regression line is known as the *residual* or *error. If the residuals are not distributed normally the simple linear regression is not applicable*. Another very important issue is that the dependent variable should be continuous and normally distributed while independent can be continuous or categorical.

Simple linear regression gave birth to multiple linear models where more than one variable related to that dependent can be assumed. As a rule the dependent variable in a *multiple linear regression* model is a continuous variable. When the dependent variable is a dichotomus variable as, for example, a disease status (presence or absence), *logistic regression* is used to consider many possible risk factors related to the disease. Regression models are one of the most frequently used techniques in modern biomedicine statistics today and few publications omit using it.

Line graphs or time series chart are mostly liked in the correlational studies and population statistics whereas the data are plotted against the time line. In multiple comparisons this form of data presentations can help in finding the relations and trade-off between the variables. Time is always plotted on the horizontal axis, and data on the vertical axis. Main economic characteristics of Lithuania, in particular GDP growth, unemployment rate and inflation are plotted on line graphs in Figure 10.9.

This completes the presentation of the pictorial tools commonly used in describing, summarizing and visualizing the data in cross sectional studies.

Chapter 11. Core principles of systematic reviews

Systematic review studies

A **systematic review** is a literature review focused on a research question that tries to identify, appraise, select and synthesize all high quality research evidence relevant to that question. The methodology of systematic review is to minimize the bias associated with single studies and nonsystematic reviews.

Key words:
Systematic review
Meta-analysis
Narrative reviews

Systematic reviews of high-quality randomized controlled trials are crucial to evidence-based medicine. Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. An understanding of systematic reviews and how to implement them in practice is becoming mandatory for all professionals involved in the delivery of health care. Besides health interventions, systematic reviews may concern clinical tests, public health interventions, adverse effects, and economic evaluations (University of York. NHS Centre for Reviews & Dissemination 2009).[14] Systematic reviews can help provide many types of information useful to policymakers, including information about the nature and extent of a problem, and the potential benefits, harms, uncertainties, and costs of interventions and policies. Policymakers may also want to know about the impact on different groups in various settings. Systematic reviews can also help answer questions about how best to disseminate information and innovations; about a particular community's receptiveness to proposed interventions—whether the interventions are appropriate to local culture and context; and about the factors influencing study outcomes. Systematic reviews are not limited to medicine and are quite common in other sciences such as psychology, nursing, occupational therapy, physical therapy, public health, educational research, sociology and business management.

Systematic review for epidemiologic questions sum up the results of primary scientific studies that meet explicit criteria. They provide an overview of current scientific literature through a definable and rigorous method in which available studies themselves are the units of analysis.

14 http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf

According to the Dictionary of Epidemiology systematic reviews are defined as "the application of strategies that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on a specific topic". The process involves specifying research questions, finding relevant literature, assessing the quality of the studies, comparing and contrasting the results of the individual studies, and interpreting the results. Sometimes the terms *systematic review* and *meta-analysis* are used interchangeably, but it is more correct to view the meta-analysis as a subset of a systematic review. Meta-analysis may be, but is not necessarily, used as part of this process. **Meta-analysis** is a statistical analysis of results from separate studies, examining sources of differences in results among studies, and leading to a quantitative summary of the results if the results are judged sufficiently similar to support such synthesis (Greenhalgh 1997a, Greenhalgh 1997b). In the biomedical sciences, meta-analysis is the systematic, organized, and structured evaluation of a problem of interest, using information (commonly in the form of statistical tables or other data) from a number of independent studies of the problem. A frequent application is the pooling of results from a set of randomized controlled trials, which in aggregate have more statistical power to detect differences at conventional levels of statistical significance (Porta et al. 2008).

Scientific knowledge builds on what has been done before. Summing up current knowledge including identification of important gaps is therefore an essential scientific activity. There are several key approaches to summing up existing literature: narrative review, systematic review, including meta-analysis.

Individual studies
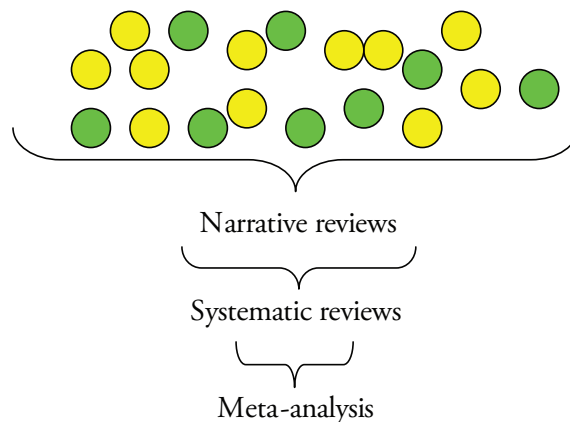
Narrative reviews summaries of what we know about a given problem. They are a good tool for providing background knowledge on a research question. Narrative reviews can give panoramic view of the issue and provide key information on relevant concepts or theory. To survey the literature is the best way to get background information on a topic. However, narrative reviews can suffer from unconscious biases.

Systematic reviews can reduce bias and enhance transparency. A systematic review attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. The process of conducting a systematic review is explicit enough that another researcher could replicate the process and results. Systematic reviews are scientific studies and have own methodology. The Cochrane database provides the tools for systematic reviews and meta-analysis and serves for disseminating the results of systematic reviews in biomedicine[15]. A systematic review must be carefully designed in order to avoid the possibility of biases and errors that may affect the results.

The key characteristics of a systematic review are (Higgins, Green & Cochrane Collaboration 2008):

- a clearly stated set of objectives with pre-defined eligibility criteria for studies;
- an explicit, reproducible methodology;
- a systematic search that attempts to identify all studies that would meet the eligibility criteria;
- an assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias; and
- a systematic presentation, and synthesis, of the characteristics and findings of the included studies;

It is essential that systematic reviews be undertaken by more than one person. This ensures that tasks such as selection of studies for eligibility and data extraction can be performed by at least two people independently, increasing the likelihood that errors are detected. In additional, involving a diverse team provides broader perspectives for specifying research questions and for interpreting the data and that enhances the methods and strengthens the results. Having specialists in a subject matter area will help to ensure that the questions asked are consistent with current theory and practice. Review teams must include expertise in the topic area being reviewed and include, or have access to, expertise in systematic review methodology (including statistical expertise). In addition to the team who will undertake the review there may be a advisory groups who are consulted at various stages, including, for example, health care professionals, patient representatives, service users and experts in research methods.

Explicit set of methods should be outlined in a protocol. Decisions about the review question, inclusion criteria, search strategy, study selection, data extraction, quality assessment, data synthesis and plans for dissemination should be addressed. Specifying the methods in advance reduces the risk of introducing bias into the review. For example, a clear inclusion criterion avoids selecting studies according to whether their results reflect a favored conclusion (University of York. NHS Centre for Reviews & Dissemination 2009).

Protocol should include background, objectives, methods (criteria for selection studies, types of studies, types of participants, types of intervention, types of outcome measures), search methods for identification of studies, data collection and analysis, references, supplementary information.

15 www.cochrane.org/resources/handbook

One of the most important issues in the planning stage is framing a research question. Systematic reviews should set clear questions, the answers to which will provide meaningful information that can be used to guide decision-making. The review question can be framed in terms of the population, intervention(s), comparator(s) and outcomes of the studies that will be included in the review. These elements of the review question, together with study design, will then be refined in order to determine the specific inclusion criteria that will be used when selecting studies for the review (University of York. NHS Centre for Reviews & Dissemination 2009).

Some groups formulate final research questions by repeated rounds of writing and revising potential questions, other begin with diagrams (logic models) that outline the hypothesized etiologic process the intervention is attempting to address. The groups draw steps between causes, interventions, and outcomes for populations or individual persons. Research questions are constructed from the populations, interventions, comparison exposures and outcomes included in diagrams. Defining and selecting interventions for review is another challenge. The challenge is to establish criteria that define a reasonably homogenous and relevant group of interventions, programs, policies and target populations (Brownson, Petitti 2006).

The next step is to identify and obtain the relevant research. Studies should be included or excluded based on explicit criteria related to characteristics of the intervention, comparison exposure and outcome of interest. The inclusion criteria should capture all studies of interest. If the criteria are too narrowly defined there is a risk of missing potentially relevant studies and the generalizability of the results may be reduced. On the other hand, if the criteria are too broad the review may contain information which is hard to compare and synthesize. Inclusion criteria also need to be practical to apply; if they are too detailed, screening may become overly complicated and time consuming (University of York. NHS Centre for Reviews & Dissemination 2009).

Studies are not always published as full papers in peer-reviewed journals. Studies not readily available in the published literature are sometimes called "gray" literature. Such literature may include conference proceedings, newsletters, research reports, theses and dissertations. Finding and reviewing unpublished information may reduce publication bias by including negative or neutral studies that are less likely to be published.

Many reviews limit by the language. If reviews include only studies reported in English, their results and inferences may be biased. Whenever feasible, all relevant studies should be included regardless of language. However, realistically this is not always possible due to a lack of time, resources and facilities for translation.

The protocol should specify the databases and additional sources that will be searched, and also the likely search terms to be used. A narrow search using more specific terms may yield a smaller set of articles to review, a broader search using less specific terms will yield more articles, a greater proportion of which may be out of scope.

Study selection is usually conducted in two stages: an initial screening of titles and abstracts against the inclusion criteria to identify potentially relevant papers followed by screening of the full

papers identified as possibly relevant in the initial screening. The protocol should specify the process by which decisions on the selection of studies will be made. This should include the number of researchers who will screen titles and abstracts and then full papers, and the method for resolving disagreements about study eligibility (University of York. NHS Centre for Reviews & Dissemination 2009). Critical appraisal of the included studies is an essential part of a systematic review. A review should be based on the best quality evidence available. Methodological quality has been defined as the extent to which a study's design, conduct and analysis have minimized selection, measurement, and confounding biases. Medicine and epidemiology have traditionally identified the randomized controlled trial (RCT) as the gold standard for assessing intervention efficacy. It should not be assumed that all studies of the same basic design (e.g. RCT) are equally well-conducted. Also, RCT are not always appropriate or feasible.

The process of collecting data from studies for a systematic review is known as data abstraction. The protocol should outline the information that will be extracted from studies identified for inclusion in the review and provide details of any software to be used for recording the data.

As far as possible, the protocol should specify the strategy for data synthesis. The most common approach to presenting data in systemic reviews is to use evidence tables, which are tabular representations of key features of the study designs and important results. We present an example on reporting the systematic review on effects of hepatitis B immunization for newborn infants of hepatitis B surface antigen-positive mothers[16].

Resources to learn more about systematic reviews

1. EPPI Centre. An Institute of Education center focusing on systematic reviews in education, health and social policy[17].
2. Campbell Collaboration. Independent organization producing systematic reviews on what works for education, health and social policy to build healthy and stable societies[18].
3. Cochrane Collaboration. Independent organizations producing systematic reviews for health interventions[19].
4. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011[20]

---

16 http://onlinelibrary.wiley.com/o/cochrane/clsysrev/articles/CD004790/frame.html

17 http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=89

18 http://www.campbellcollaboration.org/about_us/index.php

19 http://ukcc.cochrane.org/cochrane-collaboration

20 www.cochrane-handbook.org

Chapter 12. Qualitative studies

Introduction to qualitative research
Types of qualitative research
Sampling
Data collection
Data analysis
Study validity and reliability
Comparison of quantitative and qualitative research

Introduction to qualitative research

Key words
phenomenology
grounded theory
ethnography
informant
saturation

For many years researchers have practiced mostly in quantitative studies. It happened so because social sciences (psychology, economics, sociology) and biosciences applied research methods following the way physical sciences used them. In those cases usually physical objects were analyzed, but such methods were not appropriate to evaluate complex human behavior. The interest in research in health related sciences was more focused on detection of etiological factors and less to understanding of the complexities of human behavior which are sometimes dismissed by quantitative approaches. Besides that, the researchers in health area were usually trained to conduct quantitative studies and lacked knowledge in carving out qualitative studies.

Consequently last decades enlightened the way for the expanded use of qualitative studies, which became more and more attractive to the scientists.

What is qualitative research?

Qualitative researchers aim to gather an in-depth understanding of human behavior and the reasons that govern such behavior. Here words are meaningful, not numbers. The qualitative method investigates why and how decisions are made, not "what", "where", "when", "how many". Although this kind of research is more common in social sciences, recently it took up its meaningful position in health related sciences as well.

Very often qualitative studies are a <u>prelude</u> to the quantitative studies. They are used to define the problem, generate initial hypotheses, determine the main factors and make the project for quantitative research. For instance the preliminary work undertaken for the British national survey of sexual attitudes and lifestyles could be a good example to start quantitative survey discovering the most comprehensible terms or words in common use to include in a subsequent survey questionnaire

(Pope C 2006). "In this case, face-to-face interviews were used to uncover popular ambiguities and misunderstandings in the use of a number of terms such as 'vaginal sex', 'oral sex', 'penetrative sex' and 'heterosexual'. This qualitative work had enormous value in the development of the subsequent survey questionnaire, and in ensuring the validity of the data obtained because the language in the questionnaire was clear and could be widely understood."

Also qualitative studies can be applied as a <u>reinforcement</u> and/or a <u>refinement</u> of knowledge which initially was gained in quantitative research, answering why and how things happen. For example research on cultural beliefs about hypertension has helped to explain why rates of compliance with prescribed medications vary significantly among and between white and Afro-Caribbean patients (Morgan M, Watkins C 1988).

Qualitative studies are relatively quick (to collect data) and not expensive. They are done with a small number of respondents, therefore their findings can't be generated to entire population. On the other hand qualitative studies provide a lot of valuable information which helps to go deeper into issues of interest and explore nuances related to the studied problem.

Health research has concentrated almost exclusively on analysis of causes of illness, with public health taking broader approach – causes of any status of interest. Quantitative studies were rather appropriate methods to work out many matters. However there is now a growing interest in processes beyond this.

Toni Faltermaier discusses quantitative and qualitative research in etiology (Faltermaier T 1997).

"Many studies on a variety of physical and psychic disorders followed; in sum, they demonstrated significant correlations between the rates of prevalence or incidence and social indicators such as social class, gender, and urban/rural housing. For example, evidence of a negative correlation between social class and the prevalence of psychotic disorders suggested that the living conditions of lower social classes contributed to their higher risk of developing schizophrenia. However, as correlative associations cannot unequivocally be interpreted as causal, the results of epidemiological studies are not conclusive concerning the question of etiology. A clear decision between the hypotheses of social causation or social selection is not possible relying solely on the results of an epidemiological association.

Thus, etiological research changed in a way so as to include more specific psychosocial influences and more individual processes. In addition, the individual is seen more as an active agent contributing to the genesis or prevention of disease by his/her risk behaviors, coping efforts, appraisals or preventive health behaviors. Methodologically, this change in the research subjects, resulting in a complex array and interaction of causal factors at different levels, has important consequences. First, there are new demands for data collection. These psychosocial and individual variables can be studied only by including subjective reports. The persons studied and their perceptions of life conditions must receive more attention in health research if these new questions are to be answered. The personal meaning of events, the context in which they occur, the efforts individuals make to cope with life stressors or to prevent an illness they feel at risk from – data on these topics demand sensible and less structured methods of data collection in order to provide information not anticipated by the researchers. For those purposes standardized questionnaires are not adequate research tools."

Types of qualitative research

There is a variety of qualitative research designs and strategies. Most authors discuss the main three as phenomenology, grounded theory, ethnography.

Phenomenology. The aim is to show how complex meanings are built out of simple units of direct experience. The aim of phenomenology is to investigate phenomena, which might be known, but the main point is to disclose the nature and ways of human experience. Phenomena may be events, situations, experiences or concepts. The researcher studies daily life, how the people experience it, and what those experiences mean. The experiences of different people are bracketed, analyzed and compared to identify the essences of the phenomenon. For example concerning phenomenon back pain, correlation studies tell about what people experience back pain, how many and the apparent causes. Randomized controlled trials of drugs compare the effectiveness of one analgesics against others. But what is it actually like to live with back pain? What are the effects on peoples' lives? What problems does it cause? A phenomenological study might explore, for example, the effect that back pain has on sufferers' relationships with other people by describing the strain it can cause in marriages or the effect on children of having a disabled parent (Hancock B 1998).

Phenomenological research will not necessarily provide definitive explanations but it will raise awareness and increases insight.

Grounded theory was originated with Glaser and Strauss' and their work on interactions between health care professionals and dying patients. The goal of this type of qualitative study is to derive inductively from the data a theory that is „grounded" in the data. Grounded theory lets researcher to search for theoretical concepts according initial material and build substantive theory, which is distinguished from grand or formal theory. There are no hypotheses starting the study. The researcher gives a subjective sense to gathered data (texts), interprets ideas of informants (the respondent in qualitative research) through his/her understanding.

The main feature is the development of new theory through the collection and analysis of data about a phenomenon. It goes beyond phenomenology because the explanations that emerge are completely new knowledge and are used to develop new theories about a phenomenon. In health care settings, the new theories can be applied enabling us to approach existing problems in a new way. For example, our approaches to health promotion or the provision of care (Hancock B 1998).

Grounded theory examines the "six Cs" of social processes (causes, contexts, contingencies, consequences, covariances, and conditions) to understand the patterns and relationships among these elements (Strauss, Corbin 1998).

Ethnography (*gr. etnos – people, grapho – to write*). This type of qualitative research was developed by anthropologists specifically to study cultures and people. Of course they should have some common parameter as geographical, religious, tribal, shared experience. The aim of ethnography is to describe, interpret and point out how people accumulate their experience during

their individual life or in group/organization/community, how they recognize their social group, its traditions and symbols, social environment, system, relations with different cultures. Therefore the common sense to apply ethnography is to elucidate the specifics of groups in the same culture, rather than look for the disparities between different cultures.

In health care settings, researchers may choose an ethnographic approach because the cultural factor is suspected to affect the population's response to care or treatment. For example, cultural rules about contact between males and females may contribute to reluctance of women from an Asian subgroup to take up cervical screening. Ethnography helps health care professionals to develop cultural awareness and sensitivity and enhances the provision and quality of care for people from all cultures (Hancock B 1998).

For a qualitative study to be an ethnographic, a sociocultural interpretation of the data must be done.

Sometimes other types of qualitative research are applied: basic interpretive, case study, narrative analysis, critical qualitative research, postmodern research.

Basic interpretive. The researcher is interested in understanding how participants make meaning of a situation or phenomenon. A rich descriptive account of the findings is presented and discussed, using references to the literature that framed the study in the first place.

Case study. It is an intensive in-depth description and analysis of phenomenon or social unit such as individual, group, institution or community.

Narrative analysis. The key is the use of stories as data as biography, autobiography, life story, oral story, autoethnography and life narratives. Narrative analysis usually reflects perspective of the teller rather than that of society. Context is very important, as narratives can take many forms in different audiences and due to other factors. Discourse analysis is one of actual analyses of narratives. It examines the written text of the story for its component parts or assesses the spoken words by looking for intonation, pauses. Sequence and consequence are particularly important in narrative analysis

Critical qualitative research focuses less on individuals than on context.

Postmodern research report does not follow a specific format, each has its own rhythm and structure (Merriam 2002).

Those types of research are not the same, although have some characteristics in common. The terms as „grounded theory", „ethnography", „narrative analysis" and so on can not be used interchangeably.

Sampling

Planning a qualitative study follows the same common rules as in quantative research. However there are some disparities.

After the problem statement the next step is to select a sample from which the data will be collected.

What sample?

Qualitative research follows a different logic and random sampling makes little sense here. Instead, it uses theoretical or non-probability samples for selecting a population for study, the so called purposive samples. The sample is not intended to be statistically representative. Here the individual participants are selected deliberately for their specific characteristics that are important to the study. The main point is to get explicit content, because information-rich cases are appropriate for the research, as they can give most about issues that need to answer study questions. For example, a study investigating the experiences of vegetarians will select people who have practiced this life style for a determined number of years.

How many?

Qualitative samples are usually small in size. According to Ritchie, Lewis and Elam (2003), there are three main reasons for this (Ritchie J, Lewis J 2003).

"First if the data are properly analyzed, there will come a point where very little new evidence is obtained from each additional fieldwork unit. This is because a phenomenon needs only to appear once to be part of the analytical map. There is therefore a point of diminishing return where increasing the sample size no longer contributes new evidence.

Second, statements about incidence or prevalence are not the concern of qualitative research. There is therefore, no requirement to ensure that the sample is of sufficient scale to provide estimates, or to determine statistically significant discriminatory variables.

Third, the type of information that qualitative studies yield is rich in detail. There will therefore be many hundreds of ‚bites‘ of information from each unit of data collection. In order to do justice to these, sample sizes need to be kept to a reasonably small scale‘

Sample size depends on the scope of the study, the type of study, data collection method, the quality of the data, the data source and other criteria, specific for particular study. Recommendations for sample size vary in different sources, therefore it‘s researcher‘s decision to chose the size that could uncover the study problem".

Depending on a study extent there might be ten, twenty, fifty informants. In fact, although the sample size is predicted, it might be decided to end up in the process, when the saturation (when no new information is gained with new informants) is reached.

Data collection

The data are collected gradually. Here data collection is done simultaneously with data analysis. Having data/information from the first interview, first observation or first document, the material is analysed, adjustments for data collection are done. The following meetings, observations or document reading can be changed to get the best information to answer the study question.

There are three major sources of gathering data for a qualitative research studies: <u>interviews, observations and documents</u>. Sometimes only one method is used, sometimes they are combined to answer the research question. For example, in studying how some intervention in health institution gave an effect, you might interview patients and staff, observe the work day of different personnel, review available documents reflecting study question, etc. "If at all possible, researchers are encouraged to use more than one method of data collection as multiple methods enhance the validity and findings" (Merriam 2002).

<u>Interviews</u>, the primary data collection strategy in qualitative research, range from highly structured to unstructured. The latter most of all allows interviewee to feel free in his talk. Besides that interviews can be single and group interviews (consensus panel, focus groups, natural group or community). From group interviews the focus group is the most commonly used strategy in qualitative research. In different types of qualitative research data collecting method yields different result. For instance the interviews in grounded theory and phenomenology are to elicit participant's story, in discourse analysis – to capture participant's language (it is not assumed that the researcher and participant mean the same when they use the same words).

<u>Observation</u> represents a direct encounter with the phenomenon of interest comparing to the information obtained during an interview. Observer can be active or passive, known or unknown. This technique of data collection is useful when "activity, event or situation can be observed firsthand, when a fresh perspective is desired, or when participants are not able or willing to discuss the phenomenon under the study" (Merriam 2002).

<u>Documents</u> can be written, oral, visual (such as photographs) or cultural artifacts. There also might be researcher-generated documents (which are prepared after the study has begun – participants might be asked to keep a diary, take pictures, write life story, etc.).

Data analysis

The collection of qualitative data frequently results in big amount of information. Qualitative analysis is not governed by codified rules in the same way as quantitative data analysis. A. Bryman presents main steps of analytic induction (Bryman A 2008):

Analysis in qualitative research distills content textual data to a set of categories or concepts from which the final issue can be gained.

Although the idea of analysis is the same, there are some disparities in each type of research.

„For example phenomenological analysis is primarily a writing exercise, as it is through the process of writing and rewriting that the researcher can distill meaning. Analysts use writing to compose a story that captures the important elements of the lived experience. By the end of the story the reader should feel that she has vicariously experienced the phenomenon under study and should be able to envision herself (or someone else who has been through the experience) coming to similar conclusions about what it means.

The objective of a discourse analysis is to understand what people are doing with their language in a given situation. Thus, the coding phase for a discourse analysis entails identifying themes and roles as signified through language use.

Grounded theory involves a constant comparison method of coding and analyzing data through three stages: open coding (examining, comparing, conceptualizing, and categorizing data); axial coding (reassembling data into groupings based on relationships and patterns within and among the categories identified in the data); and selective coding (identifying and describing the central phenomenon, or "core category," in the data). Ideally, each interview or observation is coded before the next is conducted so that new information can be incorporated into subsequent encounters. Themes identified through the coding of initial interviews may also be explored in follow-up interviews." (Starks, Trinidad 2007).

In ethnographic research the researcher attempts to interpret data from the perspective of the population under study. "The results are expressed as though they were being expressed by the subjects themselves, often using local language and terminology to describe phenomena. For example, a researcher may explore behavior which we traditionally in the westernized medical world would describe as mental illness. However, within the population under study, the behavior may not be characterized as illness but as something else – as evidence that the individual is "blessed" or "gifted" in some way." (Hancock B 1998).

There are also computer assisted qualitative data analysis software, such as Nvivo Research Software, NUD*IST, ATLAS.ti or Etnograph. They are beneficial when the sample size is bigger and can give the main structuring and coding. However they neither spontaneously classify, nor compare data. Accordingly, the analytical researcher's mind is the most appropriate tool.

There is no standard form how to present findings, thus a lot of disparities in that sense are found. It depends on the researcher and auditorium for which the results are presented.
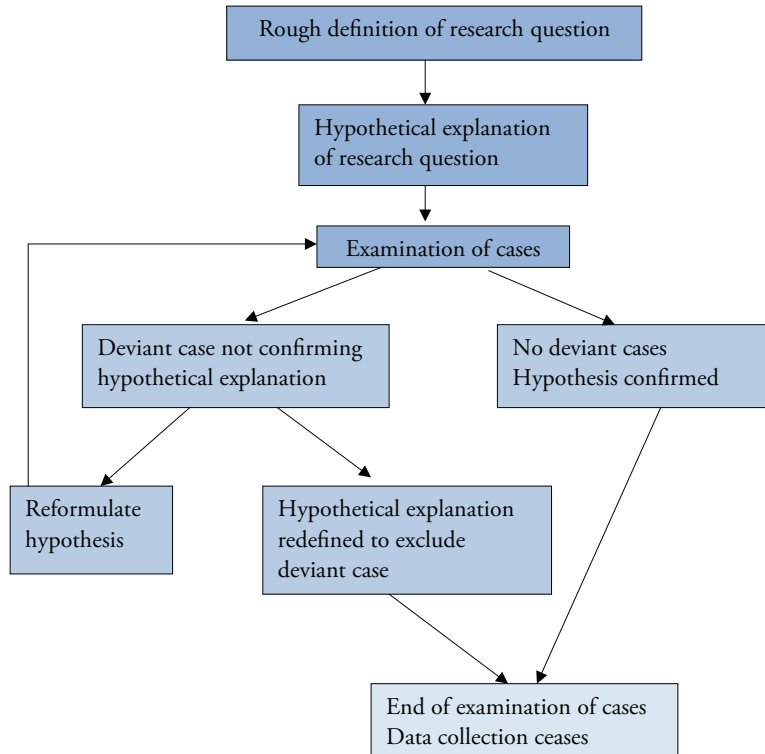
Study validity and reliability

The most important point in study validity and reliability is if the study question is appropriate for qualitative inquiry. The interest lies not in superficial opinions or in cause and effect, but in-depth understanding of phenomenon, an individual or a situation.

Merriam Sh. and associates 2002 presents **strategies** for promoting **validity** and **reliability** in qualitative research (Merriam 2002).

- **T**riangulation – pooled technique which helps to judge the reality. Using multiple investigators, sources of data or data collection methods to confirm emerging findings
- Member checks – taking data and tentative interpretations back to the people from whom they were derived and asking if they were plausible
- Peer review/examination – discussions with colleagues regarding the process of study, the congruency of emerging findings with the raw data and tentative interpretations
- Researcher's position or reflexivity – critical self-reflection by the researcher regarding assumptions, worldview, biases, theoretical orientation, and relationship to the study that may affect the investigation
- Adequate engagement in data collection – adequate time spent collecting data such that the data become "saturated"; this may involve seeking discrepant or negative cases of the phenomenon
- Maximum variation – purposefully seeking variation or diversity in sample selection to allow for a greater range of application of the findings by consumers of the research
- Audit trail – a detailed account of the methods, procedures, and decision points in carrying the study
- Rich, thick descriptions – providing enough description to contextualize the study such that readers will be able to determine the extent to which their situation matches the research context, and hence, whether findings can be transferred.

External validity and generalizability. In qualitative research it is not possible to generalize statistically. Generalizability here needs to be thought differently from quantitative research, i.e. what we learn in particular situation we can transfer to similar situations subsequently encountered. Case-to-case or user generalizability is common practice in medicine, where the practitioner decides whether a previous case is applicable to the present situation. Thus the researcher must provide enough details of the study context so that comparison can be made.

Validity and reliability depend very much on the ethics of the researcher.

Comparison of features of qualitative and quantitative research

	Qualitative	Quantitative
Start	Researcher may only know roughly in advance what he/she is looking for. Methods are selected on a way.	Researcher knows clearly in advance what he/she is looking for. Methods are clearly stated before the start.
The aim	To give a complete, detailed description.	To classify features, count them, and construct statistical models in an attempt to explain what is observed.
Theory	Emergent	Testing
Questions asked	Why? How ? In what way? In connection with what other behavior or thoughts?	How many? How much? Where? How often?
Design	The design emerges as the study unfolds.	All aspects of the study are carefully designed before data are collected.
Findings	Develop hypotheses, gain insights, explore language options, and refine concepts. Inductive (theory is an outcome of research)	Test hypotheses, prioritize factors, provide data for mathematical modeling and projection. Deductive (theory guides research))
Data	Data are in the form of <u>words</u>, pictures or objects. They are rich, deep.	Data are in the form of <u>numbers</u> and statistics. They are hard, reliable.
Data gathering instrument	Researcher is the data gathering instrument.	Researcher uses tools, such as questionnaires or equipment to collect numerical data.
Data gathering manner	Process	Static
Role of interviewer	**Critical**: interviewer must think and frame questions and probes in response to whatever respondents say. A highly trained professional is advisable.	**Important**, but interviewers need only to be able to read scripts. They should not improvise. Minimally trained, responsible employees are suitable.
Questioning development in data collections	Questions/observation "guidelines" vary in order and phrasing from group to group and from interview to interview. New questions are added, old ones dropped. Usually unstructured	Order and phasing of questions carefully controlled. Should be (ideally) exactly the same for each interview. Structured.
Number of respondents / informants	Fewer tending to last a longer time	Many in order to give a projectable scientific sample

	Qualitative	**Quantitative**
Results	Subjective, individual's interpretation of events is important.	Objective, seeks precise measurement and analysis of target concepts.
	Qualitative data are more 'rich', analysis is time consuming, and less able to be generalized.	Quantitative data are more efficient, able to test hypotheses, but may miss contextual detail.
Researcher's role	Researcher tends to become subjectively immersed in the subject matter. Researcher is close, expresses his point of view.	Researcher tends to remain objectively separated from the subject matter. Researcher is distant, expresses point of view of participants.
Conclusions	Contextual understanding	Generalization

References

1. Adami, H., Hunter, D.J. & Trichopoulos, D. 2008, *Textbook of cancer epidemiology,* 2nd edn, Oxford University Press, New York ; Oxford.

2. Bland, M. 1987, *An introduction to medical statistics,* Oxford University Press, Oxford ; New York.

3. Bowers, D. 2008, *Medical statistics from scratch: an introduction for health professionals,* 2nd edn, Wiley, Chichester.

4. Brownson, R.C. & Petitti, D.B. 2006, *Applied epidemiology: theory to practice,* 2nd edn, Oxford University Press, New York.

5. Bryman A 2008, *Social research methods,* 3rd ed. edn, Oxford University Press Inc.;, New York.

6. Corbett, S.J., Rubin, G.L., Curry, G.K. & Kleinbaum, D.G. 1993, „The health effects of swimming at Sydney beaches. The Sydney Beach Users Study Advisory Group", *American Journal of Public Health,* vol. 83, no. 12, pp. 1701–1706.

7. Dadoniene, J., Grazuleviciute, E., Rugiene, R., Luksiene, A., Stropuviene, S. & Jurgelenas, A. 2007, „Rheumatoid arthritis in Lithuania: need for external help from the onset of disease", *Disability and rehabilitation,* vol. 29, no. 18, pp. 1470–1474.

8. Dodson, T.B. 2007, „A guide for preparing a patient-oriented research manuscript", *Oral surgery, oral medicine, oral pathology, oral radiology, and endodontics,* vol. 104, no. 3, pp. 307–315.

9. Doll, R., Peto, R., Boreham, J. & Sutherland, I. 2004, „Mortality in relation to smoking: 50 years' observations on male British doctors", *BMJ (Clinical research ed.),* vol. 328, no. 7455, pp. 1519.

10. Faltermaier T 1997, „Why public health research needs qualitative approaches. Subjects and methods in change.", *European Journal of Public Health,* vol. 7, no. 4, pp. 357–363.

11. Fletcher, R.H. & Fletcher, S.W. 2005a, *Clinical epidemiology: the essentials,* 4th edn, Lippincott Williams & Wilkins, Philadelphia ; London.

12. Fletcher, R.H. & Fletcher, S.W. 2005b, *Clinical epidemiology: the essentials,* 4th edn, Lippincott Williams & Wilkins, Philadelphia ; London.

13. Gerstman, B.B. 2003, *Epidemiology kept simple: an introduction to traditional and modern epidemiology,* 2nd edn, Wiley-Liss, Hoboken, N.J.

14. Green MD, Freedman DM, Gordis L 2000, „Reference guide on epidemiology" in *Reference Manual on Scientific Evidence*, ed. Federal Judicial Cener, pp. 333–401.

15. Greenhalgh, T. 1997a, „How to read a paper. Papers that tell you what things cost (economic analyses)", *BMJ (Clinical research ed.),* vol. 315, no. 7108, pp. 596–599.

16. Greenhalgh, T. 1997b, „Papers that summarise other papers (systematic reviews and meta-analyses)", *BMJ (Clinical research ed.),* vol. 315, no. 7109, pp. 672–675.

17. Greenland, S. 1980, „The effect of misclassification in the presence of covariates", *American Journal of Epidemiology,* vol. 112, no. 4, pp. 564–569.

18. Gregg, M.B. 2008, *Field epidemiology,* 3rd edn, Oxford University Press, Oxford.

19. Guillemin Francis, Bombardier Claire, Beaton Dorcas 1998, „Cross-Cultural adaptation of health-related quality of life measures: literature review and proposed guidlines", *J Clin Epidemiol,* vol. 46, no. 12, pp. 1417–1432.

20. Hancock B 1998, *Trent Focus for Research and Development in Primary Health Care: An Introduction to Qualitative Research.* .

21. Hennekens, C.H., Buring, J.E. & Mayrent, S.L. 1987, *Epidemiology in medicine,* 1st edn, Lippincott Williams & Wilkins, Philadelphia.

22. Hennekens, C.H., Buring, J.E., Mayrent, S.L. & Doll, R. 1987, *Epidemiology in medicine,* Little, Brown, Boston Toronto.

23. Higgins, J., Green, S. & Cochrane Collaboration 2008, *Cochrane handbook for systematic reviews of interventions,* Wiley-Blackwell, Chichester, England ; Hoboken, NJ.

24. HILL, A.B. 1965, „The Environment and Disease: Association Or Causation?", *Proceedings of the Royal Society of Medicine,* vol. 58, pp. 295–300.

25. Huck, S.W. 2004, *Reading statistics and research,* 4th edn, Allyn and Bacon, Boston, MA.

26. Hulley, S.B. 2007, *Designing clinical research,* 3rd edn, Lippincott Williams & Wilkins, Philadelphia, PA.

27. Indrayan, A. 2008, *Medical biostatistics,* 2nd edn, Chapman & Hall/CRC, Boca Raton.

28. Juul S. 2008, *Take good care of your data,* Institute of Public hHalth, Department of Epidemiology, University of Aarhus.

29. Juul S & Frydenberg M (eds) 2010, *An introduction to Stata for health researchers.*, 3rd edn, StataCorp LP, 2010.

30. Kleinbaum,D.G., Sullivan K.M., Barker N.D. 2003, *ActiveEpi. Companion textbook. A supplement for use with the ActiveEpi CD-ROM,* Springer-Verlag, New York.

31. Kurth T, S.J. 2007, „Assessment and control of confounding in trauma research", *J Trauma Stress,* vol. 20, no. 5, pp. 807–820.

32. Lang JM, Rothman KJ, Cann CI. 1998, „That confounded *P*-value", *Epidemiology,* vol. 9, no. 1, pp. 7–8.

33. Levin, K.A. 2006a, „Study design III: Cross-sectional studies", *Evidence-based dentistry,* vol. 7, no. 1, pp. 24–25.

34. Levin, K.A. 2006b, „Study design IV. Cohort studies", *Evidence-based dentistry,* vol. 7, no. 2, pp. 51–52.

35. Levin, K.A. 2006c, „Study design V. Case-control studies", *Evidence-based dentistry,* vol. 7, no. 3, pp. 83–84.

36. Levin, K.A. 2005, „Study design I", *Evidence-based dentistry,* vol. 6, no. 3, pp. 78–79.

37. Long JS (ed) 2009, *The workflow of data analysis using Stata.*, StataCorp LP, StataCorp.

38. Mann, C.J. 2003, „Observational research methods. Research design II: cohort, cross sectional, and case-control studies", *Emergency medicine journal : EMJ,* vol. 20, no. 1, pp. 54–60.

39. Merriam, S.B. 2002, *Qualitative research in practice: examples for discussion and analysis,* Jossey-Bass, San Francisco, Calif.

40. Miettinen, O.S. 1985, *Theoretical epidemiology: principles of occurrence research in medicine,* Wiley, New York.

41. Morgan M & Watkins C 1988, „Managing hypertension: beliefs and responses to medication among cultural groups. „, *Sociology of Health and Illness,* vol. 10, pp. 561–578.

42. Peacock, J. & Peacock, P.J. 2011, *Oxford handbook of medical statistics,* Oxford University Press, Oxford.

43. Peat J, B.B. 2005, *Medical statistics: a guide to data analysis and critical appraisal.* Blacwell, 2005.

44. Petrie, A. & Sabin, C. 2005, *Medical statistics at a glance,* 2nd edn, Blackwell, Malden, Mass. ; Oxford.

45. Pope C, M.N. 2006, *Qualitative research in health care,* 3rd ed. edn, Blackwell Publishing Ltd.

46. Porta, M.S., Greenland, S., Last, J.M. & International Epidemiological Association 2008, *A dictionary of epidemiology,* 5th edn, Oxford University Press, Oxford.

47. Ritchie J & Lewis J, G.E. 2003, „Designing and selecting samples." in *Qualitative research practice. A guide for social science students and researchers.*, eds. Ritchie J & Lewis J, Thousand Oaks, CA: Sage, pp. 77–108.

48. Rockhill, B., Newman, B. & Weinberg, C. 1998, „Use and misuse of population attributable fractions", *American Journal of Public Health,* vol. 88, no. 1, pp. 15–19.

49. Rothman, K.J. 1976, „Causes", *American Journal of Epidemiology,* vol. 104, no. 6, pp. 587–592.

50. Rothman, K.J. 2002, *Epidemiology: an introduction,* Oxford University Press, Oxford.

51. Rothman, K.J., Greenland, S. & Lash, T.L. 2008, *Modern epidemiology,* 3rd edn, Lippincott Williams & Wilkins, Philadelphia ; London.

52. Shapiro, S., Goldberg, J.D. & Hutchison, G.B. 1974, „Lead time in breast cancer detection and implications for periodicity of screening", *American Journal of Epidemiology,* vol. 100, no. 5, pp. 357–366.

53. Starks, H. & Trinidad, S.B. 2007, „Choose your method: a comparison of phenomenology, discourse analysis, and grounded theory", *Qualitative health research,* vol. 17, no. 10, pp. 1372–1380.

54. Strauss, A. & Corbin, J. 1998, *Basics of qualitative research: Techniques and procedures for developing grounded theory.* 2nd ed. edn, Thousand Oaks, CA: Sage.

55. Susser, M. 1991, „What is a cause and how do we know one? A grammar for pragmatic epidemiology", *American Journal of Epidemiology,* vol. 133, no. 7, pp. 635–648.

56. University of York. NHS Centre for Reviews & Dissemination 2009, *Systematic reviews: CRD's guidance for undertaking reviews in health care,* CRD, University of York, York.

57. van Duijn, C.M., Delasnerie-Laupretre, N., Masullo, C., Zerr, I., de Silva, R., Wientjens, D.P., Brandel, J.P., Weber, T., Bonavita, V., Zeidler, M., Alperovitch, A., Poser, S., Granieri, E., Hofman, A. & Will, R.G. 1998, „Case-control study of risk factors of Creutzfeldt-Jakob disease in Europe during 1993–95. European Union (EU) Collaborative Study Group of Creutzfeldt-Jakob disease (CJD)", *Lancet,* vol. 351, no. 9109, pp. 1081–1085.

58. von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gotzsche, P.C., Vandenbroucke, J.P. & STROBE Initiative 2008, „The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies", *Journal of clinical epidemiology,* vol. 61, no. 4, pp. 344–349.
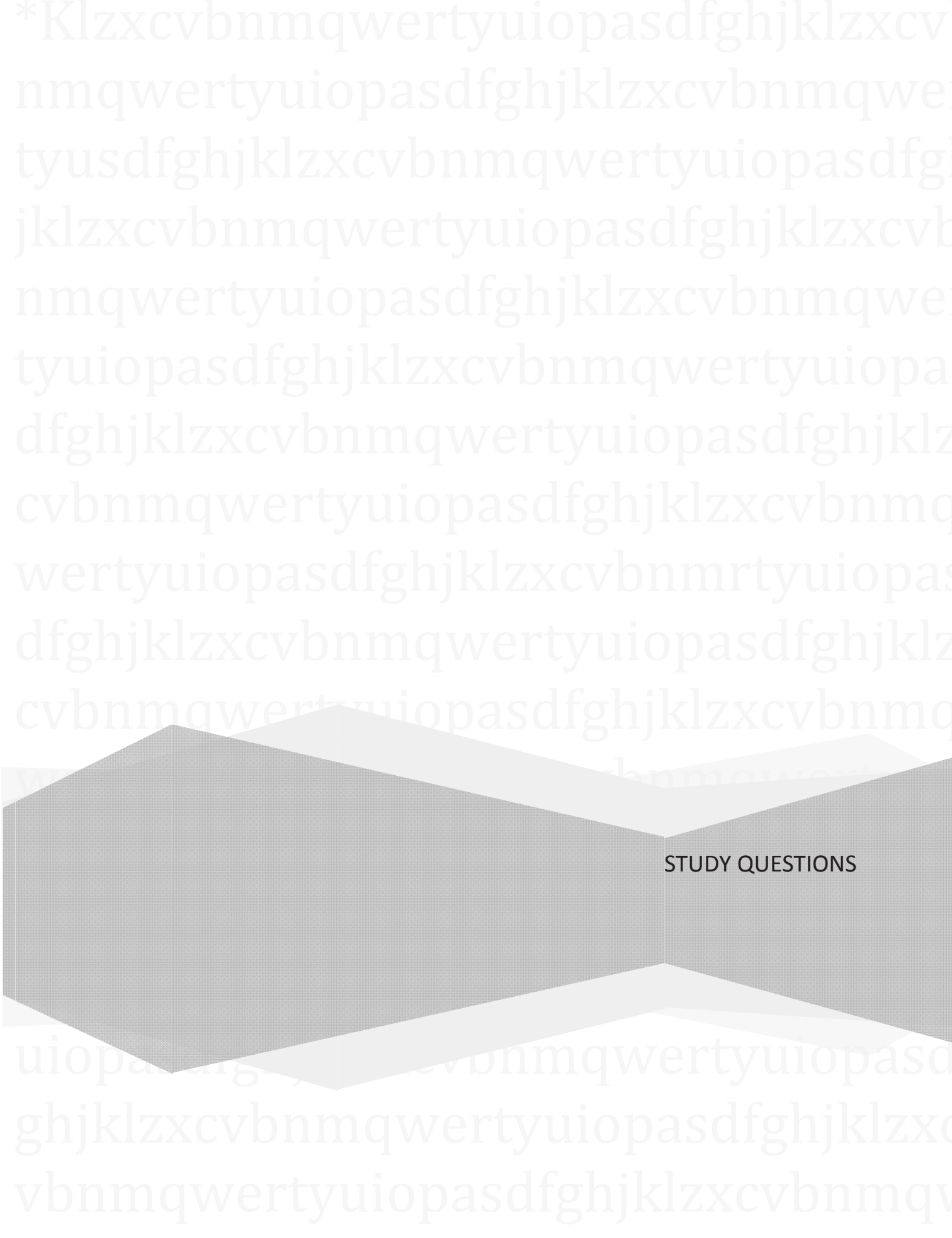
CHAPTER 1. Study Questions

1.1. Please criticize the following causal criteria and fill in the table:

Hill's Causal Criteria	Problems
1. Strength	
2. Consistency	
3. Specificity	
4. Temporality	
5. Biologic gradient	
6. Plausibility	
7. Coherence	
8. Experimental evidence	
9. Analogy	

1.2. Two measures of frequency commonly used in epidemiology, the incidence rate and the prevalence rate. In a few sentences, tell how these two measures differ.

1.3. How do we determine whether a certain disease is associated with a certain exposure?

1.4. Please fill in the table:

Relative risk (RR)	Your comments
RR=1	
RR>1	
RR<1	

1.5. When is the odds ratio a good estimate of the relative risk?

1.6. The following 2x2 table represents notation of cohort studies were the relation between a diseases and certain exposure are studied

	Disease + (number of persons)	Disease – (number of persons)
Exposed	j	k
Not exposed	m	n

Write the appropriate formula for each measure:
- A.   Incidence proportion among exposed =
- B.   Incidence proportion among not exposed =
- C.   Relative risk =
- D.   The exposure odds of the cases =
- E.   The disease odds of the not exposed =

1.7. Hypothetical cohort study:

	Myocardial infarction (MI) develops	MI does not develop	Total	Incidence per 1000 year
Smoke cigarettes	90	2910	3000	30
Do not smoke cigarettes	80	4920	5000	16

- A.   In population of smokers how much of MI that they experience is due to smoking?
- B.   How much of the MI could be prevented if they did not smoke?
- C.   What would be the total impact of a prevention program on the community (total population)?

CHAPTER 2. Study Questions

2.1. In a few sentences please define and contrast selection bias and sampling error.

2.2. Please define and contrast selection and information bias.

2.3. What is meant by confounding?

CHAPTER 3. Study questions

3.1. Having learned that Analytic studies may be either **Experimental** or **Observational** fill in the blanks with one of these two

    A.   A strength of the ??? study is the researchers control in the assignment of individuals to treatment groups.

    B.   A potential strength of the ??? study is that they are often carried out in more natural settings, so that the study population is more representative of the target population.

    C.   The randomization is used in ??? studies

    D.   Preventive clinical trial and therapeutic clinical trial both are ???

    E.   A weakness of an ??? is that randomization to treatment groups may be unethical if the comparative group will be denied a treatment which is regarded as beneficial.

    F.   ??? design is appropriate for studies of causation

3.2. What kind of study is it: **prospective cohort**; **case-control; cross sectional; case serious, clinical trial?**

    A.   Researchers administered a questionnaire to all new students at a large state university. The questionnaire included questions about behaviors such as seat belt use, exercise, smoking, and alcohol consumption. The researchers plan to distribute follow-up questionnaires at graduation and every five years thereafter, asking about health events and conditions such as diabetes and heart disease. This is an example of an **???** study.

    B.   To investigate the relationship between egg consumption and heart disease, a group of patients admitted to a hospital with myocardial infarction were questioned about their egg consumption. Another group of patients admitted to a fracture clinic were also questioned about the egg consumption using an identical protocol. The study design is ???.

    C.   To investigate the relationship between certain solvents and cancer, all employees at a factory were questioned about their exposure to an industrial solvent, and the amount and length of exposure measured. These subjects were regularly monitored, and after 10 years a copy of the death certificate for all those who died was obtained. The study design is ???

    D.   A survey was conducted of all the physicians employed at a particular hospital. Among other questions, the questionnaire asked about the number of years in study and whether or not she/he was satisfied with her career prospects. The study design is ???

    E.   A study describes the clinical course of patients who have very rare neurological disorder. Patients are identified at a referral centre that specializes in this disease. Their medical records are reviewed for patient characteristics and treatments and are then related to their current status. This study is best described as a ???

    F.   To test the efficacy of vitamin C in preventing colds, students are randomly assigned to two groups: one given 500mg of vitamin C daily, and one given placebo. Both groups are followed to determine the number and severity of subsequent colds. The study design is ???.

3.3. Is it **Preventive** or **therapeutic** trial?

A.   ??? trials are conducted on individuals with a particular disease to assess a possible cure or control for the disease. For example, we may wish to assess to what extent, if at all, a new type of chemotherapy prolongs the life of children with acute leukemia.

B.   ??? trials can be conducted on either individuals or entire populations. An example is a study in which one community was assigned (at random) to receive sodium fluoride added to the water supply, while the other continued to receive water without supplementation. This study showed significant reductions in the development of tooth decay in the community receiving fluoride.

CHAPTER 4. Study questions

4.1. In the case of the following conditions what disease frequency, **incidence** or **prevalence or both**, is more relevant to refer?

    A. Chron's disease (rare disease, long duration)
    B. Rheumatoid arthritis (not rare, with long duration)
    C. High blood pressure (common, lifelong duration)
    D. Influenza (common but seasonal occurrence, short duration)

4.2. Prevalence can be linked to the incidence with the following formula: P=IxD. The terms P, I and D in this formula represent the concepts of prevalence, incidence and duration, respectively. Rearranging the formula the disease duration can be calculated: D=P/I. Fill in the blank space for disease duration in the following table presenting the relationship among incidence, prevalence and duration of disease. (Adapted from R.H. Fletcher, S.W. Fletcher. *Clinical epidemiology. The Essentials*)

Age	Annual Incidence	Prevalence	Duration=Prevalence/Annual Incidence
0–5	6/1 000	29/1 000	
6–16	3/1 000	32/1 000	
17–44	2/1 000	26/1 000	
45–64	1/1 000	33/1 000	
65+	0	36/1 000	
Total	3/1 000	30/1 000	

    A. What does it tell about the disease character?

4.3. Data is shown for a CROSS SECTIONAL study to assess whether maternal cigarette smoking is a risk factor for low birth weight.

	Smoking mothers		Total
	Smokers	Non smokers	
Low birthweight	1,556	14,974	16,530
Normal birthweighgt	694	14,532	15,226
Total	2,250	29,506	31,756

    A. Calculate the odds ratio measures whether smoking mothers are more likely to deliver low birth weight babies.

    B. What does this odds ratio measure tell?

CHAPTER 5. Study questions

5.1. For each of the following features, choose the appropriate answer:

    A.   The investigators role regarding exposure in CASE CONTROL studies:

        a. assign   b. observe

    B.   Subject selection into groups in case control studies:

        a. self-selection   b. randomization

    C.   Directionality in cross sectional studies

        a. backwards   b. forward   c. non-directional

    D.   Timing in cross sectional studies

        a. prospective   b. retrospective   c. either

5.2. For each of the following characteristics of a study, choose the type of study a) **cross sectional**, b) **case control**, c) **cohort**:

    A.   Generating hypothesis but not testing it ???

    B.   More accurate exposure information ???

    C.   Appropriate for studying rare exposures ???

    D.   Appropriate for studying rare diseases ???

    E.   Can study multiply outcomes ???

    F.   Requires a smaller sample size ???

    G.   Can estimate risk ???

    H.   Can estimate multiple exposures and multiple outcomes at the same time ???

5.3. The relation between oral contraceptives (OCs) use and ovarian cancer was studied in a case control study and the results are given in the table.

	Use of oral contraceptives		Total
	Ever used	Never used	
Cases	93	86	179
Controls	959	683	1642
Total	1052	769	1821

    A.   Is there a chance to become ill with ovarian cancer if using OCs? What is the result and what does it mean?

CHAPTER 6. Study questions

6.1. For each of the following features, choose the option that applies to COHORT studies:

A.   The investigators role regarding exposure:
a. assignment   b. observe ???

B.   Subject selection into groups:
a. self –selection   b. randomization ???

C.   Directionality;
a. backwards   b. forwards   c. non-directional ???

D.   Timing:
a. prospective   c. retrospective   c. either ???

6.2. Fill TRUE or FALSE is the statement concerning the COHORT studies:

A.   Prospective cohort study is least prone to bias when compared with other observational design ???

B.   Cohort study can address only one outcome in the study ???

C.   Retrospective cohort study can be relatively low cost and quick ???

D.   Loss to follow up is a potential source of bias ???

E.   Prospective cohort is quite costly and time consuming ???

F.   It is suitable for rare diseases an diseases with long latency ???

G.   Prospective cohort is the only way to prove the causality of the disease ???

6.3. What is a meaning of the following statement? The 10-year risk that 45-year-old male will develop prostate cancer is 5%.

6.4. Will the 5 year risk for the same person described in the previous question be larger or smaller than 10-year risk?

6.5. The table below summarizes the results of a five-year follow up study to determine whether or not smokers who have had heart attack will reduce their risk for dying by quitting smoking. A cohort of 156 heart attack patients were studied all of whom were regular smokers up to the time of their heart attack. Seventy five continued to smoke after attack. The other 81 patients quit smoking during their recovery period. Of 75 patients that continued smoking 27 died. Of 81 who quit smoking 14 died. Here is the scheme for this study:

	Heart attack patients		Total
	Smoke	Quit	
Death	27	14	41
Survival	48	67	115
Total	75	81	156

 A. What is the directionality of the study – ???

 B. What is the timing of the study – ???

 C. The 5 year risk of dying if patient continued to smoke – ???

 D. The 5 year risk of dying if patient quit smoking – ???

 E. How much likely those who continue to smoke are to die if compared to those who quit smoking during the 5 year period (5 year relative risk) – ???

6.6. For heart attack patients, the relative risk is defined to be the risk for continuing smokers divided by the risk for smokers who quit. For the following scenarios what would be the relative risk?

 A. Continuing smokers are three time as likely to die than smokers who quit ???

 B. Continuing smokers are just likely to die as smokers who quit ???

 C. Smokers who quit are twice less likely to die than continuing smokers ???

Choices are: 0; 0.1; 0.2; 0,5; 1; 2; 3

6.7. The researchers performed a follow up study and observed an initial cohort of 1000 persons aged 65 or older for three years. Out of this cohort one hundred had lung cancer at the start of follow up, and 40 out of these one hundred died from the lung cancer. In addition, 15 persons developed cancer during the follow up period, 10 of whom died. Of the remaining 885 persons without lung cancer, 150 also died.

 A. What is the prevalence of lung cancer in the initial cohort of 1000 persons?

 B. What is the incidence (risk) of lung cancer over three – year period?

 C. What is the annual risk of lung cancer in this cohort?

 D. What is the lung cancer specific mortality risk for this cohort?

 E. What is all-cause mortality risk for this cohort?

 F. What is the case-fatality risk for the 100 lung cancer patients in the initial cohort?

	Start	3 yr follow-up	Died

Start — 3 yr follow-up — Died

1000 persons
- 100 LC ——— 40 died
- 15 LC — 10 died
- 900 no LC —
- 885 no LC ——— 150 died

Total — 200
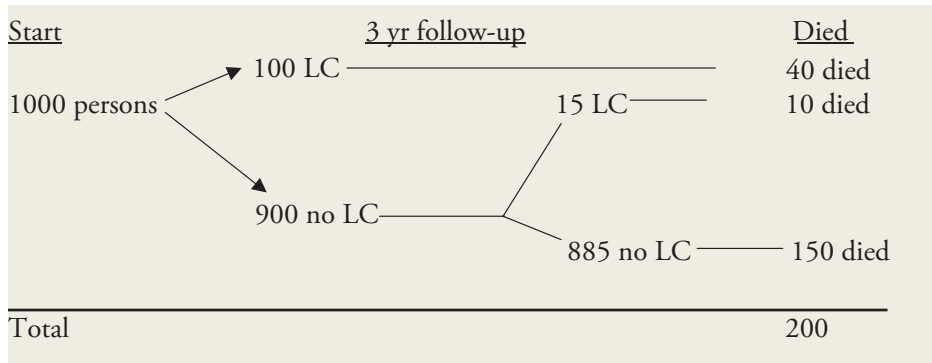
6.8. The hypothetical data in following table shows the survival time in days of 10 patients after acute myocardial infarction. The duration of the study was 15 days. Calculate the survival probabilities and build up Kaplan-Meier survival curves.

Patient nr.	Survival time in days	Outcome: died (D) or survived (S)
1	3	D
2	15	S
3	8	D
4	12	S
5	13	S
6	11	D
7	12	S
8	12	S
9	14	D
10	15	S

CHAPTER 7. Study questions

7.1. What kind of disease can be screened and suitable for developing a screening program?

7.2. What are the characteristics of a good screening test?

7.3. Read the following statements about diagnostic and screening tests and mark each one as True or False.

- A. A screening test and a diagnostic test requires a control group and a gold standard ……..........
- B. The detective method can be used to calculate the specificity of the screening test ……….........
- C. The positive predictive value of a test is an issue only with diagnostic tests ………….............
- D. When a screening program is begun, more people with disease are found on the first round of screening than on the later rounds …………..

7.4. A new screening test is introduced for pancreatic cancer. Out of 10,000 screened 800 were positive. Cancer was detected in 200 people by applying other instrumental methods (MRI and biopsy). Over the following years another 50 who had negative screening tests were diagnosed with pancreatic cancer. What is the sensitivity and specificity of this new screening test?

7.5. The idea of screening programs is to prevent disease and different ways are encountered from behavior counseling to pharmacological treatment and surgical procedures so far. Whatever the intervention, it should be effective. The rules for evaluation of effectiveness of the program are the same as for pharmacological treatment of whatever disease. Because interventions for primary prevention, like low dose aspirin, are usually given to large numbers of healthy people, they also must be very safe. For effectiveness and safety of whatever pharmaceutical product the randomized preventive clinical trial is required.

Let's consider the data from a randomized clinical trial to assess whether or not taking aspirin reduces the risk for heart disease. The exposed group received aspirin every other day whereas the comparison group received a placebo. The table of the results is shown below:

Developed heart disease	Preventive medication	
	Aspirin N=10 000	Placebo N=8 000
Yes	104	189
No	9 896	7 811

- A. What is the estimated risk to develop the heart disease for the patients in the Aspirin group?
- B. What is the estimated risk to develop the heart disease in Placebo group?
- C. The estimated risk ratio that compares the aspirin group to placebo group is?

7.6. The rheumatoid factor was determined as positive for 80 patients with rheumatoid arthritis out of 100. Equally the rheumatoid factor was found positive for 20 persons out of 1000 healthy persons.

    A.   What is the prevalence of rheumatoid arthritis in this setting?

    B.   What is the sensitivity of rheumatoid factor?

    C.   What is the specificity of rheumatoid factor?

    D.   What is positive predictive value for the person who's got positive rheumatoid factor?

    E.   What is negative predictive value for the person who's got negative rheumatoid factor?

7.7. Ankylosing spondylitis is characterized by inflammatory low back pain (ILBP). This type of back pain occurs in 95% of ankylosing spondylitis patients, but also in 10% of all healthy people. ILBP can be used as a diagnostic test.

    A.   What is the sensitivity of ILBP?

    B.   What is the specificity of ILBP?

CHAPTER 8. Study questions

8.1. Sometimes the words "hypothesis" and "theory" are used synonymously. What are the differences between "scientific hypothesis" and "scientific theory"?

8.2. What are the main purposes of the study protocol?

8.3. Can be hypothesis proven or tested?

8.4. Look at the objectives in the following abstracts. Do these objectives meet the criteria for sound epidemiologic hypotheses?

Abstract 1: Robert J. Valuck and J. Mark Ruscin. A case-control study on adverse effects: H2 blocker or proton pump inhibitor use and risk of vitamin B12 deficiency in older adults. *Journal of Clinical Epidemiology* 57 (2004) 422–428.

Objective: Acid-suppressant drugs are commonly prescribed for elderly patients, a population in which vitamin B12 deficiency is a common disorder. The purpose of this study was to examine the possible association between use of prescription histamine H-2 receptor antagonists (H2RA) or proton pump inhibitors (PPI) and vitamin B12 deficiency in older adults.

Study Design and Setting: This was a case–control study in a University-based geriatric primary care setting. Among patients aged 65 years or older with documented serum vitamin B12 studies between 1990 and 1997, 53 vitamin B12-deficient cases were compared with 212 controls for past or current use of prescription H2RA/PPI according to information in subjects' medical records.

Results: Controlling for age, gender, multivitamin use, and *Helicobacter pylori* infection, *chronic* (12 months) *current* use of H2RA/PPI was associated with a significantly increased risk of vitamin B12 deficiency (OR 4.45; 95% CI 1.47–13.34). No association was found between past or short-term current use of H2RA/PPI and vitamin B12 deficiency.

Conclusion: These findings support an association between chronic use of H2RA/PPI by older adults and development of vitaminB12 deficiency. Additional studies are needed to confirm these findings.

Abstract 2: Risk factors of gastroesophageal reflux disease methodology and first epidemiological results of the ProGERD study. M. Kulig, M. Nocon, M. Vieth, A. Leodolter, D. Jaspersen, J. Labenz, W. Meyer-Sabellek, M. Stolte, T. Lind, P. Malfertheiner and S. N. Willich *J Clin Epi*. June 2004, Volume 57, Issue 6, Pages 580–589

Objective: We describe the design and report the first results of the Progression of Gastroesophageal Reflux Disease (ProGERD) study, to our knowledge the largest prospective study of GERD patients.

Study Design and Setting: Patients were recruited at 1,253 centers in German, Austria, and Switzerland. Following an assessment of medical history, all patients were endoscoped and received esomeprazole for 2 to 8 weeks before entering the 5-year observational phase.

Results: A total of 6,215 patients (53% male, age 54 14) were included. Of these patients, 46% reported at least daily symptoms, 15% were unable to work at least once during the prior year, and

71% had visited a physician due to reflux symptoms. Barrett's esophagus (BE) was found in 11% of our GERD patients. In polychotomous regression analysis, the main factors related to the occurrence of the three GERD subgroups (nonerosive, erosive disease, and BE) were age, gender, duration of GERD, body mass index (BMI), smoking, and previous PPI use. Factors associated with longer disease duration were increasing age, male gender, BMI, increasing symptom severity, presence of erosive GERD or BE, positive family history, and smoking.

2004 Conclusion: The findings indicate that GERD is a great burden for patients, and has significant socioeconomic implications. The long term follow-up period with further endoscopic and histologic evaluations, will help further our understanding of the natural course of the disease.

CHAPTER 9. Study questions

9.1. What is the best center measure for these data?

The mean of the ten numbers: 1, 1, 1, 2, 2, 3, 5, 8, 12, 17

9.2. *Researcher 1* conducts a clinical trial to test a drug for a certain medical condition on 30 patients all having that condition. The patients are randomly assigned to either the drug or a look-alike placebo (15 each). Neither patients nor medical personnel know which patient takes which drug. Treatment is exactly the same for both groups, except for whether the drug or placebo is used. The hypothesis test has null hypothesis "proportion improving on the drug is the same as proportion improving on the placebo" and alternate hypothesis "proportion improving on the drug is greater than proportion improving on the placebo." The resulting *p*-value is $p = 0.15$.

Researcher 2 does another clinical trial on the same drug, with the same placebo, and everything else the same except that 200 patients are randomized to the treatments, with 100 in each group. The same hypothesis test is conducted with the new data, and the resulting *p*-value is $p = 0.03$. Are these results contradictory?

9.3. Let's assume that you compared two means and obtained a *P* value equal to 0.03.

Which are the correct definitions of this *P* value?

A.   There is a 3% chance of observing a difference as large as you observed even if the two population means are identical (the null hypothesis is true).

B.   Random sampling from identical populations would lead to a difference smaller than you observed in 97% of experiments, and larger than you observed in 3% of experiments.

C.   There is a 97% chance that the difference you observed reflects a real difference between populations, and a 3% chance that the difference is due to chance.

9.4. Please comment or **criticize the next statements**:

A.   The p value is the probability that the null hypothesis is incorrect.

B.   P<0.05 has an objective basis.

C.   Rejections of H_0 are infallible.

D.   Small P values provide unassailable support for a causal theory.

E.   Statistical "significance" implies practical importance.

CHAPTER 10. Study questions

10.1. Construct a scatter plot to visualize the relationship between the weight and height of ten men.

Height, cm	Weight, kg
172	76
176	79
179	80
180	81
180	82
182	83
184	93
185	90
170	69
167	70

10.2. Are the variables normally distributed?

10.3. What kind of relationship exists between the two variables and how can you prove it?

10.4. What is the equation of this relationship?

10.5. What do the coefficients mean?

CHAPTER 11. Study questions

11.1. What are the advantages of systematic reviews?

11.2 What makes a good systematic review?

11.3. Why everything with "review" in the title is not a systematic review?

CHAPTER 12. Study questions

By: Hancock B. Trent Focus for Research and Development in Primary Health Care: An Introduction to Qualitative Research. Trent Focus, 1998

Exercise 1

Look at the research projects listed below. In which projects would you expect to see a qualitative approach used and in which projects would you expect to see a quantitative approach? Why?

A.   A comparison of the effectiveness of drug A versus drug B in the treatment of migraine.

B.   An exploration of the role of the Practice Manager in the primary health care team: a study of four practices.

C.   A descriptive study of school nurses' experiences of dealing with boys who have eating disorders.

D.   A national survey of patients' knowledge of the causes of heart disease.

Exercise 2

Consider the following list of research problems and consider what would be the most appropriate qualitative research method for each one. If you think that more than one method would be appropriate, explain why.

A.   The role of Specialist Nurses in community care

B.   Developing a primary health care service for the Chinese population in one city

C.   What is advocacy in primary health care?

D.   An evaluation of the Polyclinic – a one stop primary health care centre

Exercise 3

It is not possible to demonstrate the complete procedure of content analysis within the confined space available in this pack. However, exercise 4 provides an opportunity to look at an excerpt from a transcript and begin the process of categorizing data.

The following text is an excerpt from the transcript of an interview conducted by a community psychiatric nurse with a woman following discharge from hospital. The excerpt deals with the woman's recollection of being admitted and how she felt at that time.

Read the transcript carefully and complete the following tasks.

1.   Make a note of all the items of data you consider to be potentially interesting.

2.   Identify "categories" of data.

3.   How many categories have you identified?

4.   Do some items of data potentially relate to more than one category?

5.   Can you identify major and minor categories?

Interviewer: What were your first impressions when you were first admitted to hospital

Respondent: Its hard to remember. I was so terrified. I didn't know what to expect. I was so ashamed that I was going to a loony bin. I thought everybody would be mad. But then I saw Ann. I knew her and at first I couldn't believe it, she's not mad, why is she here? Then she came up to me and smiled and said hello and she started asking me about Bill and the kids then she asked me if I was visiting someone and I told her "No, I've come in" and she told me why she was here. She didn't seem to think it was strange at all.

Interviewer: Who's Ann?

Respondent: She used to live next door to me at my last house before we moved.

Interviewer: So was it better when you saw Ann?

Respondent: Yes. Well, yes and no. It was good to see someone I knew but I didn't know what to think about it all. I mean, she was in there and I had no idea. Looking back a little while afterwards I realised that just because you go into a psychiatric hospital it doesn't mean you're mad. I wasn't and I knew she wasn't. Well, I hadn't thought so.

Interviewer: So before you arrived at the hospital, is that what you thought? That it would be full of mad people.
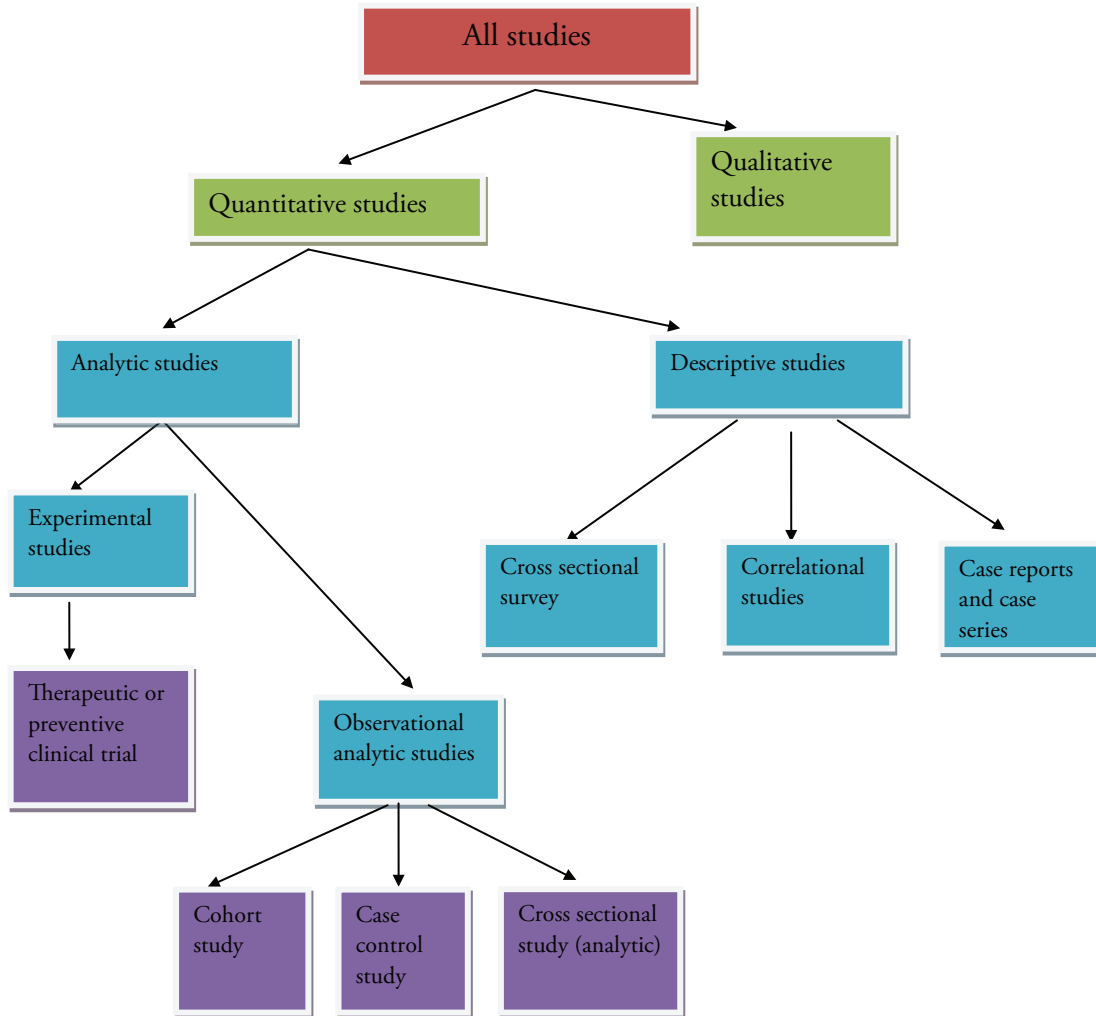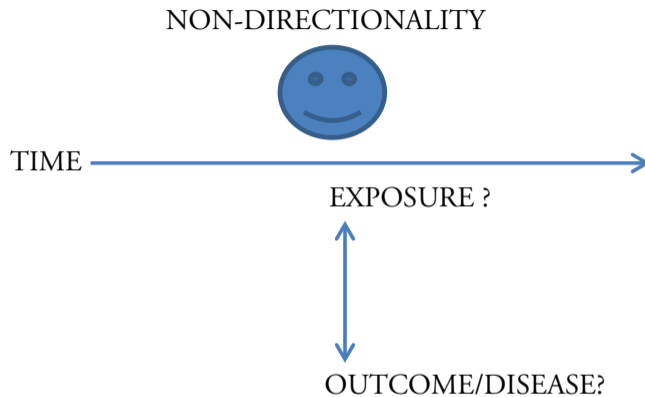
Figure 3.1 Classification tree of different types of studies.

Figure 4.1 Scheme of a cross sectional study design.
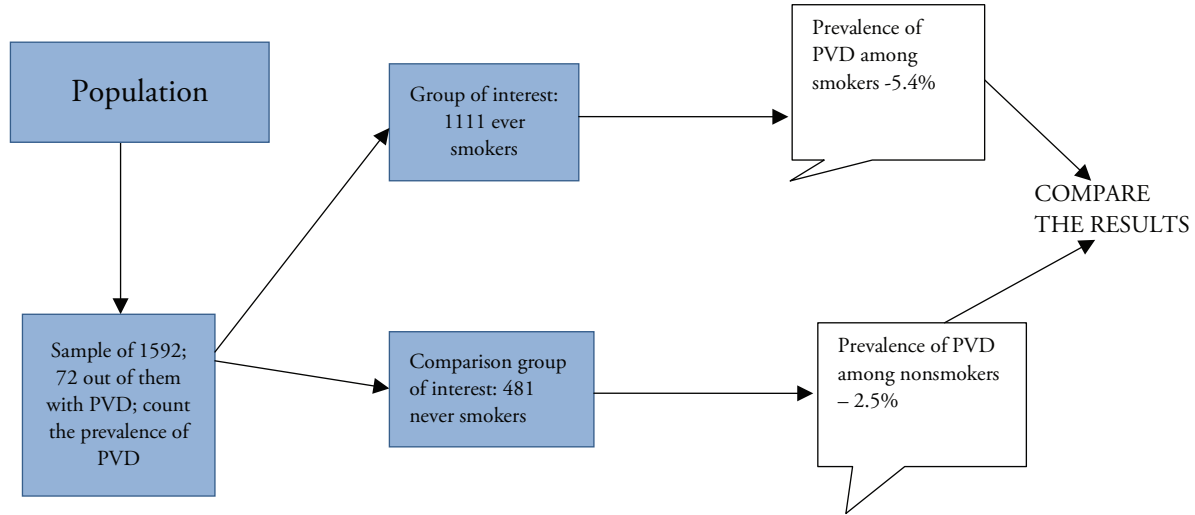
Figure 4.2 Scheme of a cross sectional study of peripheral vascular disease (PVD) among Scotish population.

BACKWARDS   DIRECTIONALITY
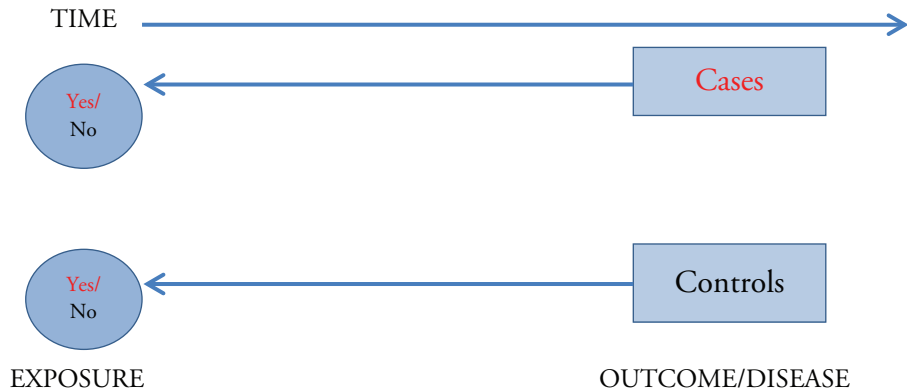
Retrospective   STUDY BEGINS

TIME

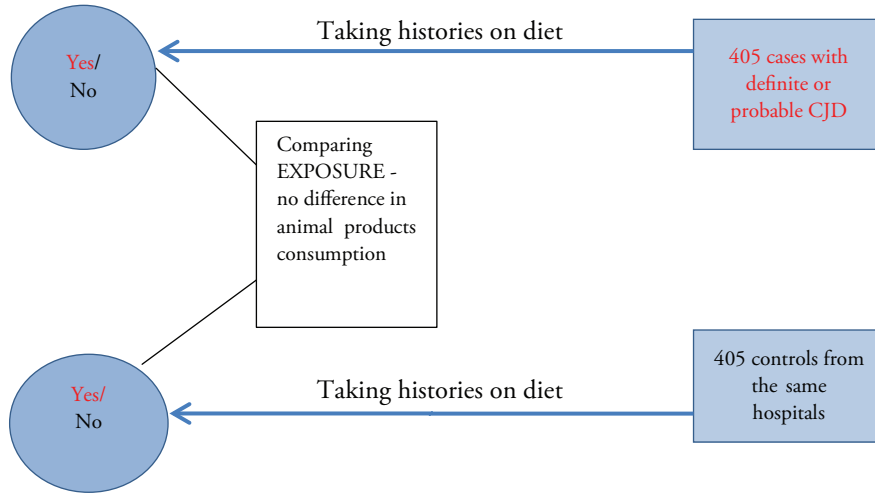Yes/No   Cases

Yes/No   Controls

EXPOSURE   OUTCOME/DISEASE

Figure 5.1 Scheme of a case control study.

Figure 5.2 Scheme of a case control study of Creuzfeld – Jacob Disease in European Union
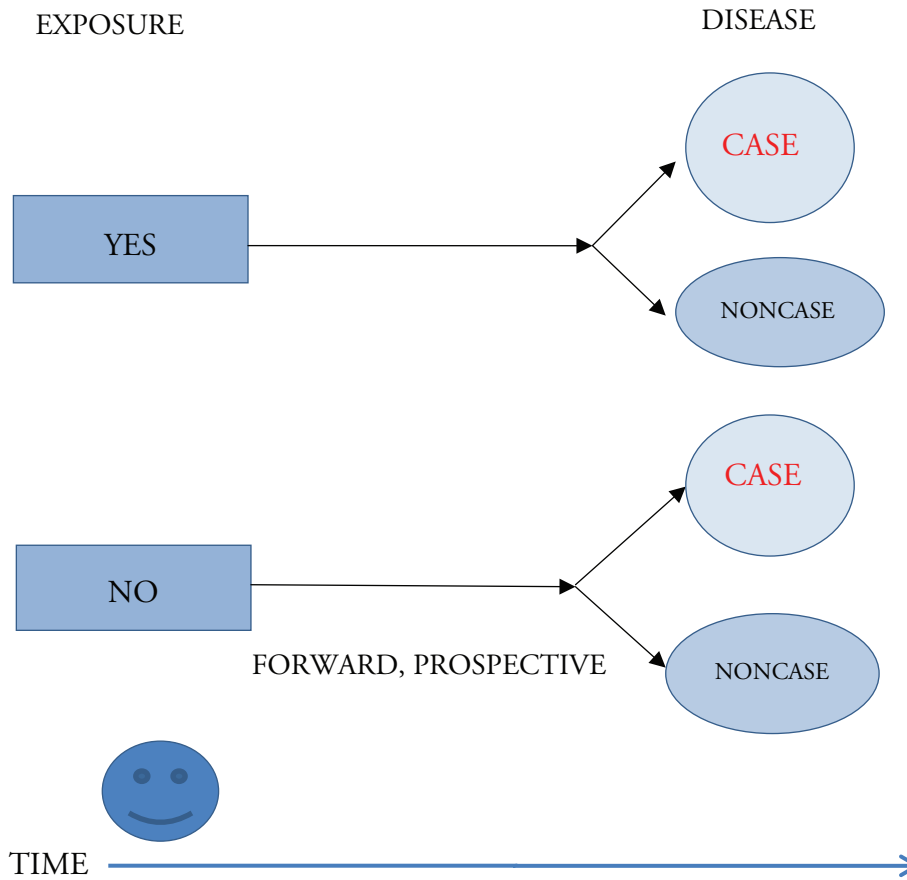
Figure 6.1 A prospective cohort study.
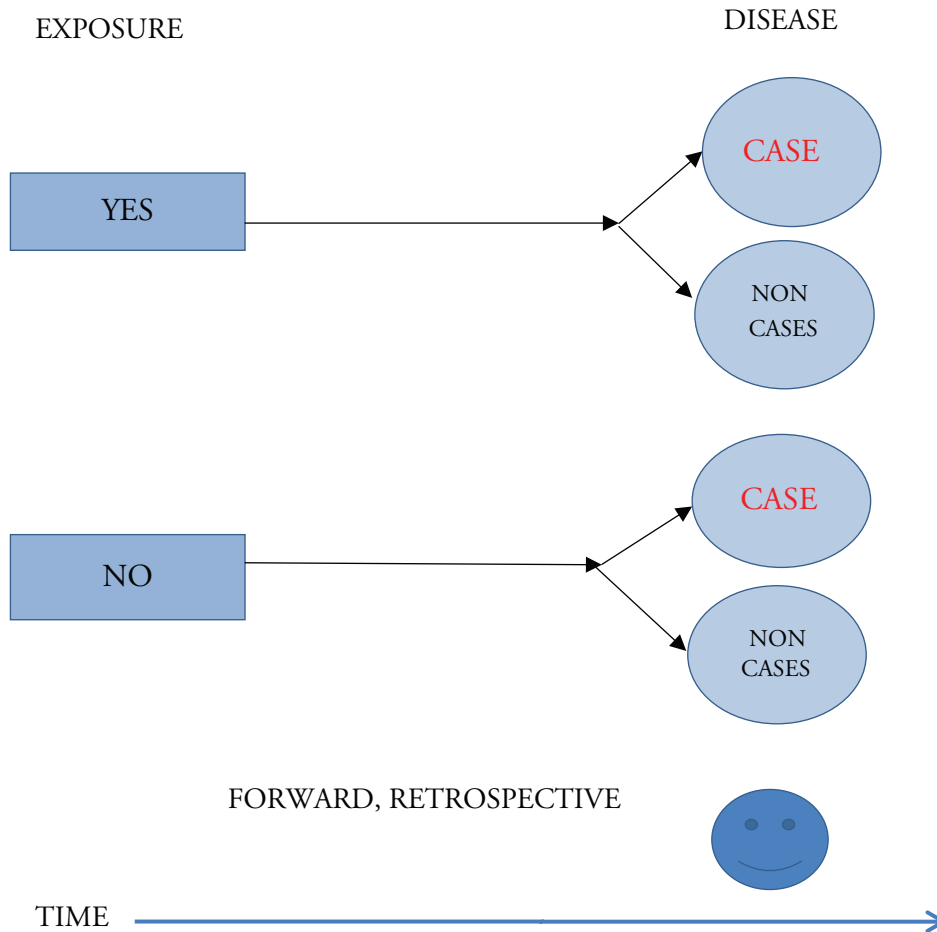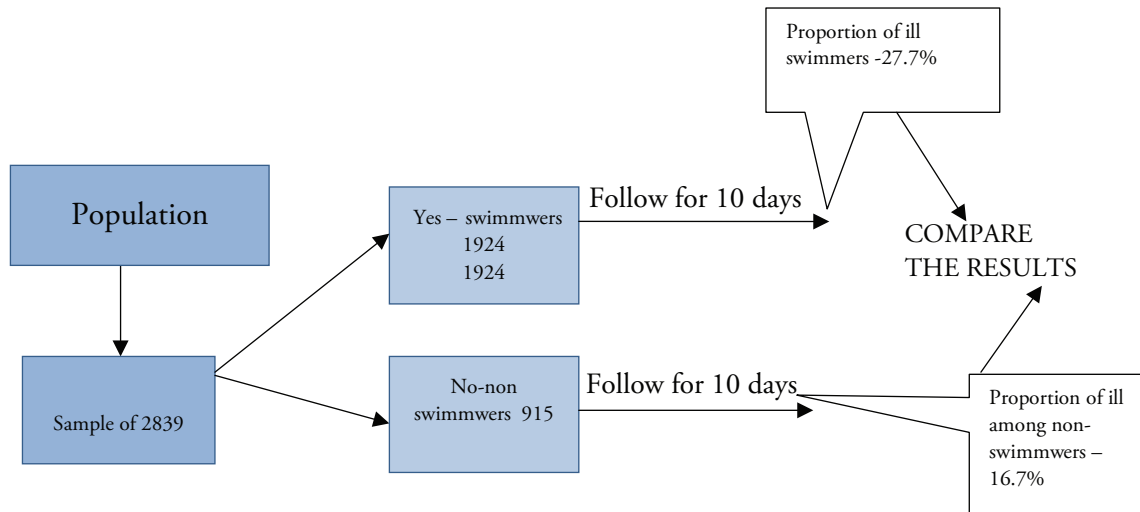
Figure 6.2. A retrospective cohort study.

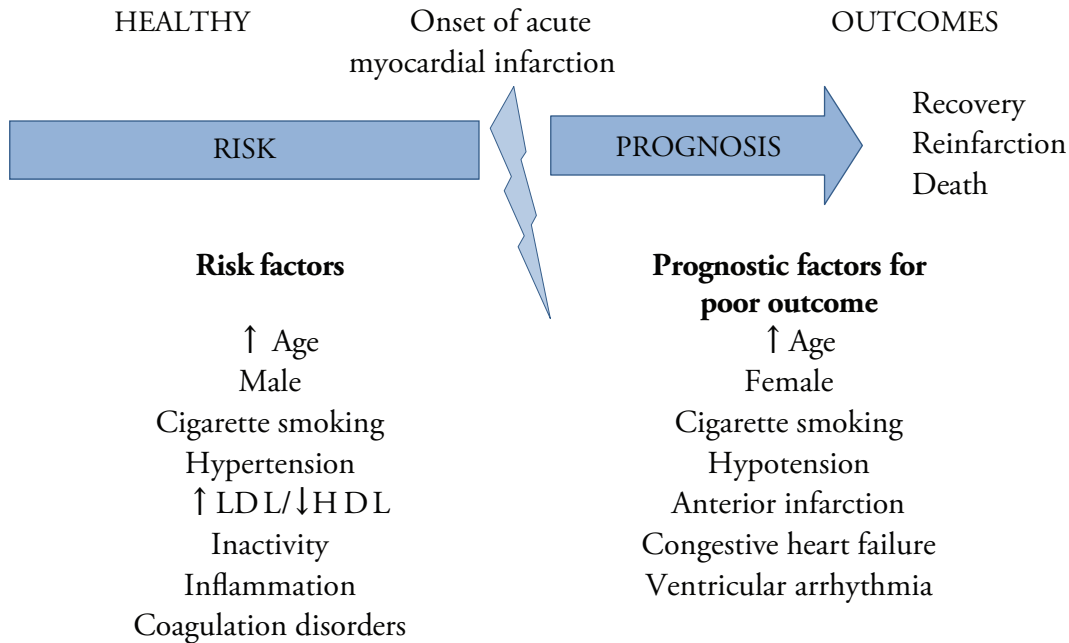**Figure 6.3** A prospective forward cohort study of Sydney Beach users.

HEALTHY   Onset of acute   OUTCOMES
                 myocardial infarction

RISK   PROGNOSIS

                                          Recovery
                                          Reinfarction
                                          Death

Risk factors   **Prognostic factors for
                                  poor outcome**

↑ Age   ↑ Age
Male   Female
Cigarette smoking   Cigarette smoking
Hypertension   Hypotension
↑ L D L/↓H D L   Anterior infarction
Inactivity   Congestive heart failure
Inflammation   Ventricular arrhythmia
Coagulation disorders

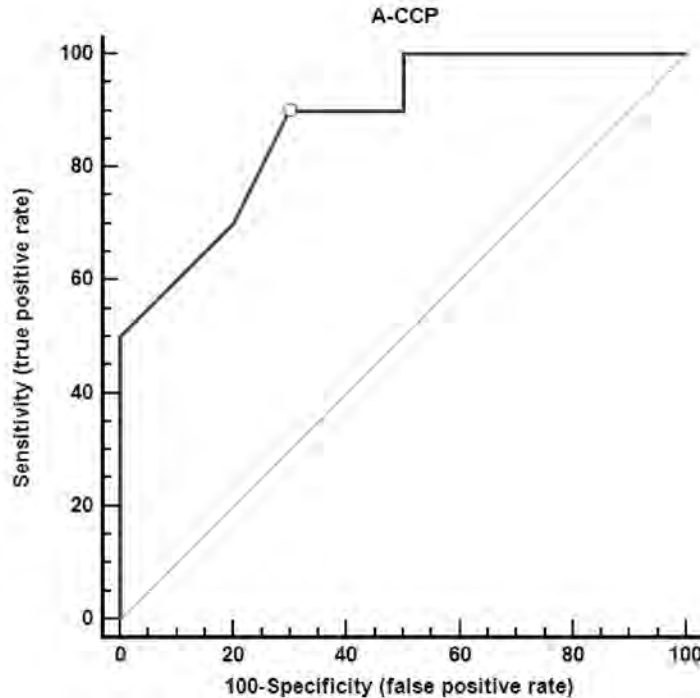**Figure 6.4** Risk and prognostic factors for acute myocardial infarction.
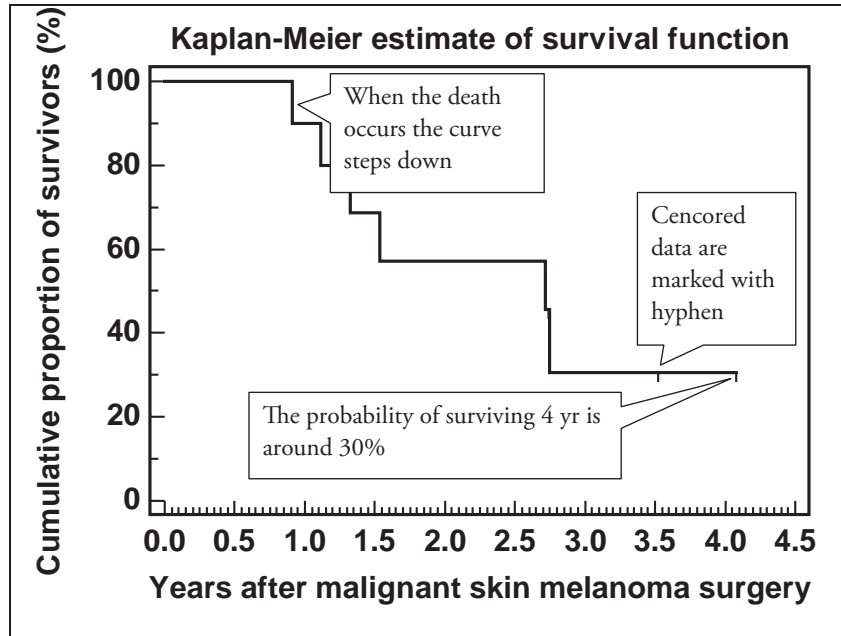
Figure 6.5 Kaplan Meier curve demonstrating the survival function after malignant skin melanoma surgery.

Figure 7.1 Visual illustration of the relationship between sensitivity and specificity for the diagnostic test capability to diagnose rheumatoid arthritis. The dot on the upper left corner illustrates the specificity of 70% and sensitivity of 90% for the value of 12 optical units of the antibodies against citrulinated proteins.
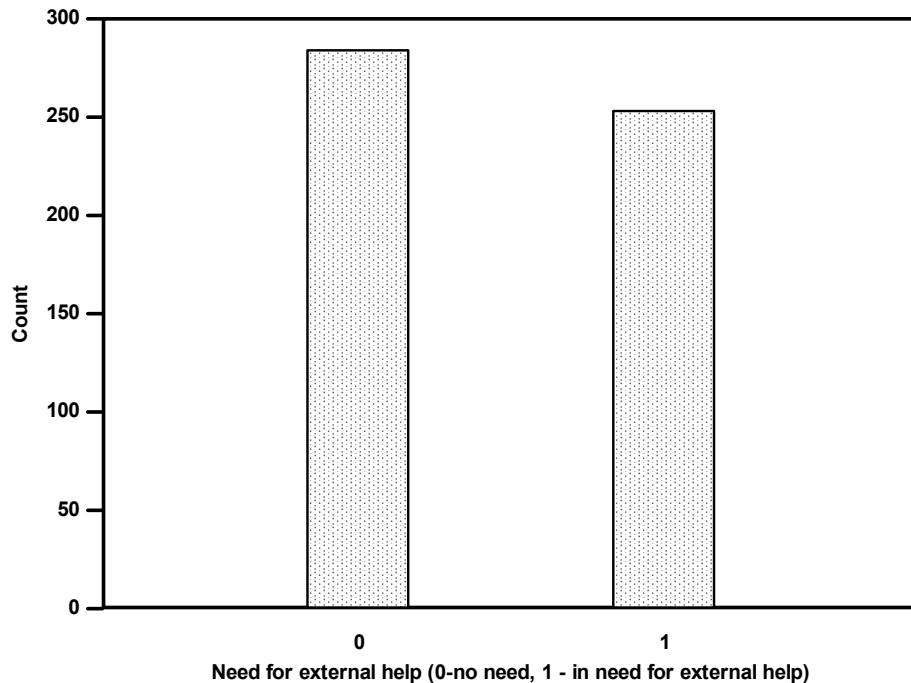
Figure 10.1. The simple bar chart for 537 rheumatoid arthritis patients divided by the need of external help. Counts displayed on vertical axis and two categories on horizontal axis.
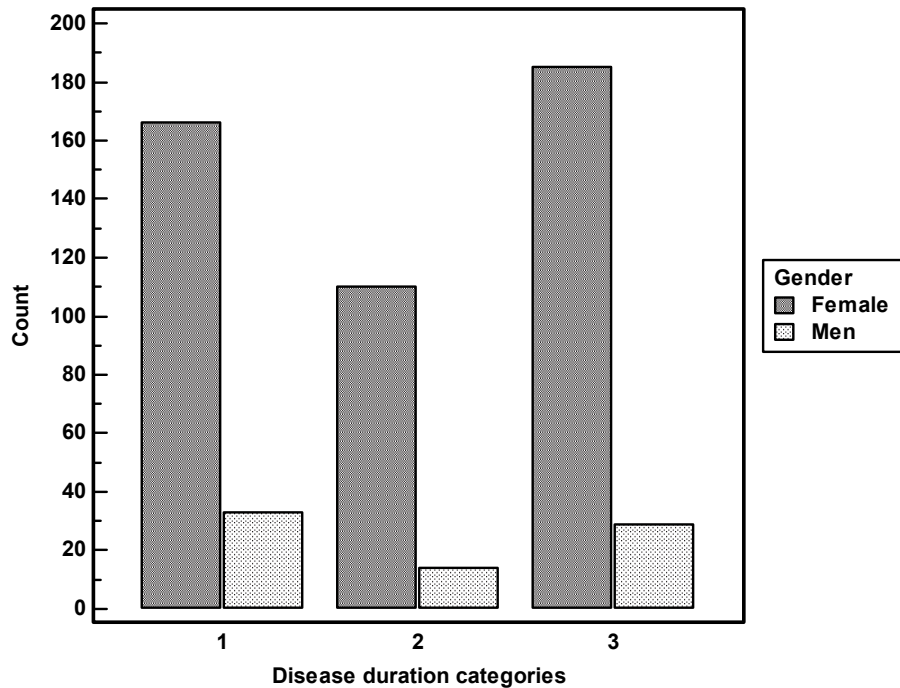
Figure 10.2 The clustered bar chart for the number of rheumatoid arthritis patients divided by gender in three different disease duration categories (1 – up to 5 yr; 2 – 6–10 yr; 3 – more than 10 yr).
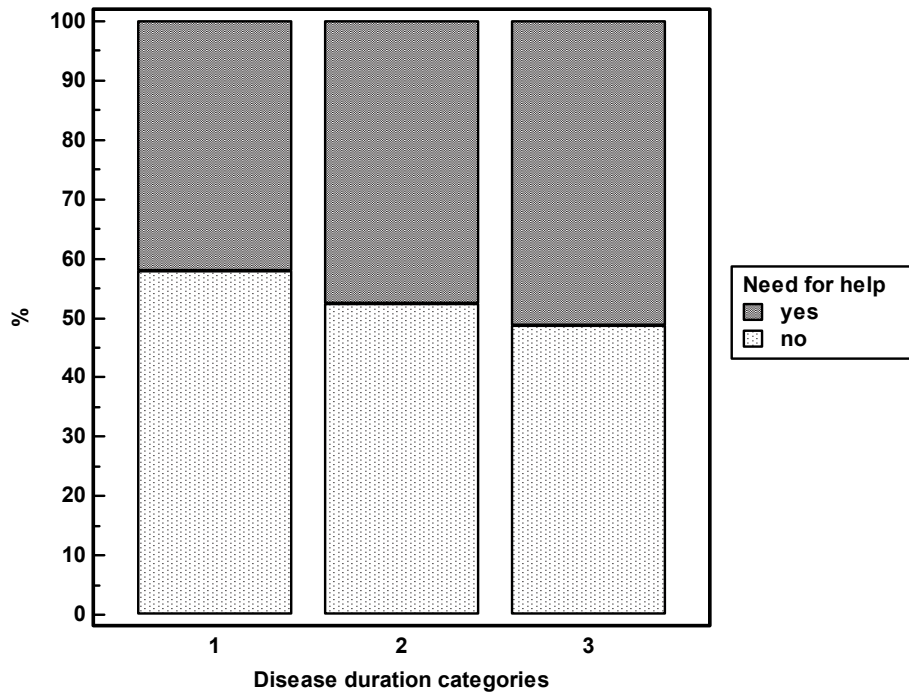
Figure 10.3 100% stacked bar chart for the proportion of rheumatoid arthritis patients in need for external help in three different disease duration categories (**1** – up to 5 yr; **2** – 6–10 yr; **3** – more than 10 yr).
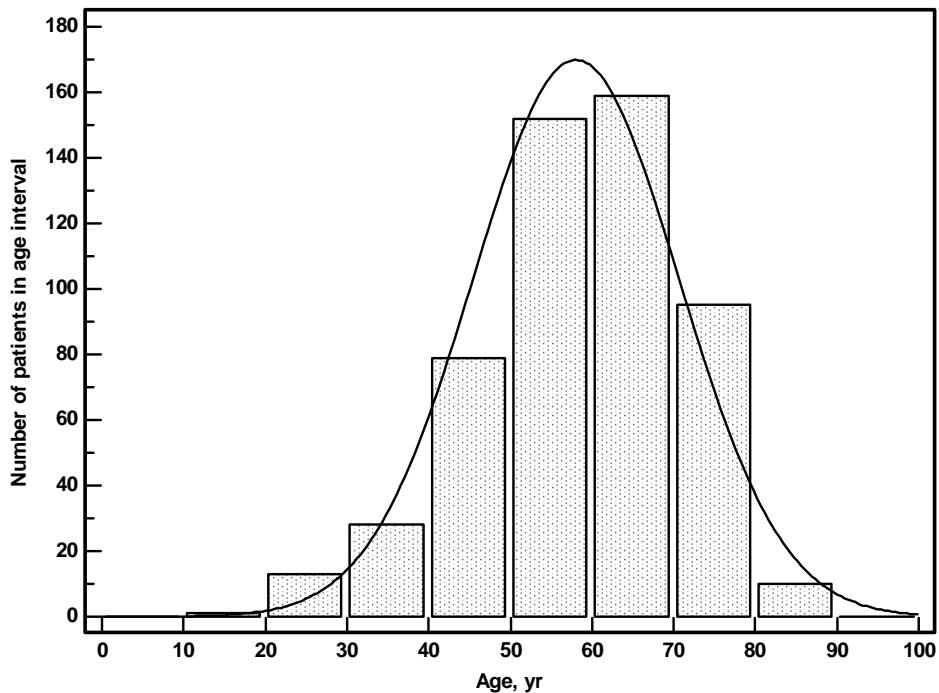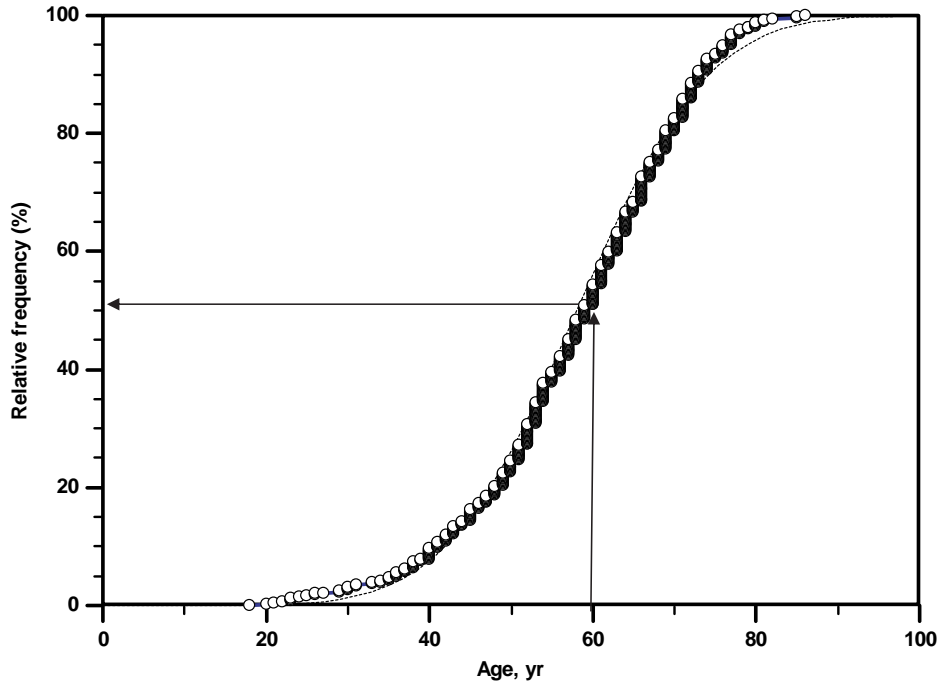
Figure 10.4 The histogram of age distribution among 537 patients with rheumatoid arthritis. The histograms are often used to check if the data fit the normal distribution.

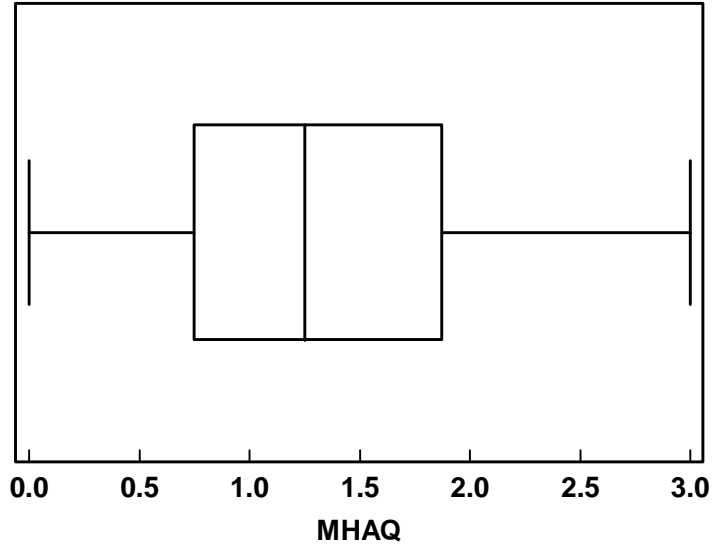**Figure 10.5** The relative cumulative frequency curve for the percentage of cumulative age data.

Figure 10.6 The box plot of modified health assessment evaluated by questionnaire (MHAQ). Explanations are presented in the text.
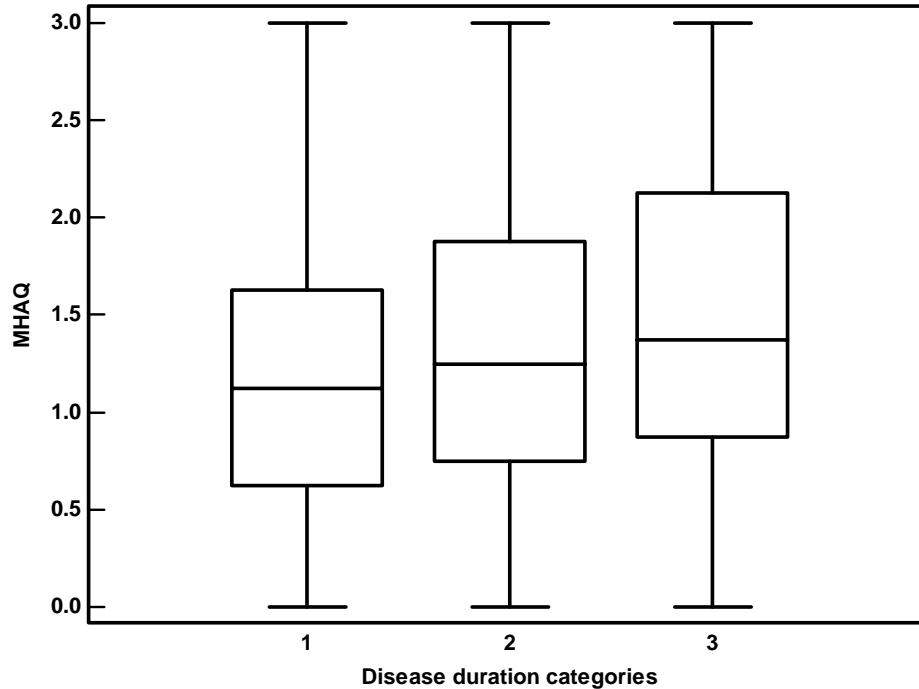
Figure 10.7 The box plots of modified health assessment (MHAQ) in three different disease duration categories (**1** – up to 5 yr; **2** – 6–10 yr; **3** – more than 10 yr).
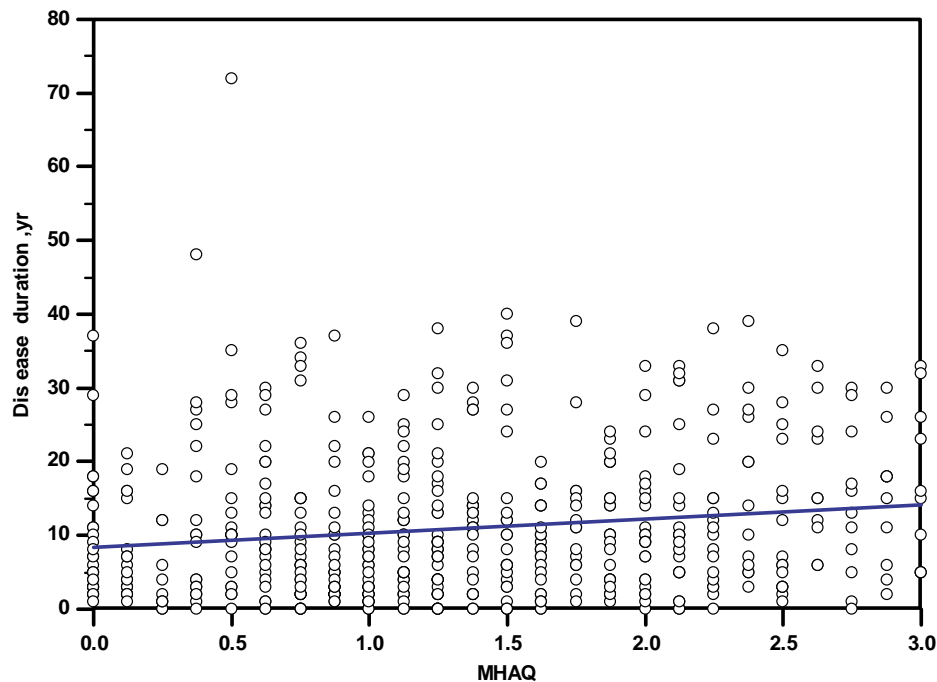
Figure 10.8 The scatterplot and regression line for relationship between modified health assessment (MHAQ) and disease duration.
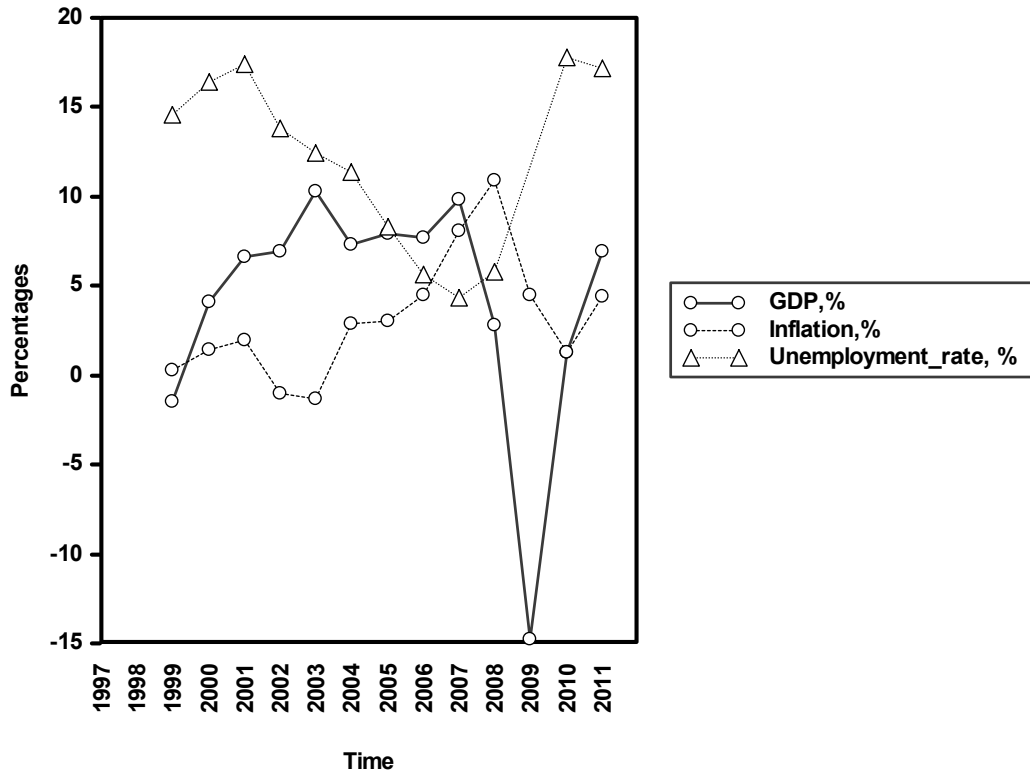
Figure 10.9 The main characteristics of Lithuania's economy since 1997. The GDP depression in 2009 is clearly visible.

CHAPTER 1. Answers

Study Questions:

1.1.

Hill's Causal Criteria	Problems (See: Rothman KJ, Greenland S, Lash TL, 2008)
1. Strength	Unmeasured confounder might be responsible for the strong association. Weak association can also reflect causality. Each component cause in a given sufficient cause has the same etiologic significance. Over a span of time, the strength of the effect of a given factor on disease occurrence may change because the prevalence of its causal complement in various mechanisms may also change, even if the causal mechanisms in which the factor and its cofactors act remain unchanged. Strong association is neither necessary nor sufficient for causality, and weakness is neither necessary nor sufficient for absence of causality.
2. Consistency	Lack of consistency, does not rule out a causal association, because some effects are produced by their causes only under unusual circumstances. Consistency serves only to rule out hypotheses that the association is attributable to some factor that varies across studies.
3. Specificity	Exposure can be associated not only with single disease.
4. Temporality	If an exposure causes disease, the exposure must occur before the disease develops. This criterion is inarguable.
5. Biologic gradient	Biologic gradient refers to the presence of a dose–response or exposure–response curve with an expected shape. A threshold may not be reached because exposure was not long enough or intense enough. Also, the threshold could be quite low and above the threshold little extra risk occurs. Some biological gradients are not monotonic, but have a peak and then decline
6. Plausibility	Plausibility can change with the times.
7. Coherence	Absence of coherent information should not be taken as evidence against an association being considered causal. The presence of conflicting information may indeed refute a hypothesis, but one must always remember that the conflicting information may be mistaken or misinterpreted.
8. Experimental evidence	Evidence from human experiments, however, is seldom available for epidemiologic research questions, and animal evidence relates to different species and usually to levels of exposure very different from those that humans experience. Uncertainty in extrapolations from animals to humans often dominates the uncertainty of quantitative risk assessments.
9. Analogy	Analogy provides a source of more elaborate hypotheses about the associations under study; absence of such analogies reflects only lack of imagination or experience, not falsity of the hypothesis.

1.2.

The prevalence rate is the ratio of the number of cases to the population. The incidence rate is the ratio of the number of *new* cases *in a period of time* to the remaining population at risk

1.3.

To determine whether a certain disease is associated with certain exposure we must determine using data obtained in case-control or cohort studies, whether there is an excess risk of the disease in persons who have been exposed to certain agent. Errors are inevitable any epidemiological study, even in the best conducted randomized trial. Thus, when interpreting findings from an epidemiological study, it is essential to consider how much the observed association between an exposure and an outcome may have been affected by errors in the design, conduct and analysis.

1.4.

Relative risk (RR)	Your comments
RR=1	Risk in exposed equal to risk in unexposed, no saaociation
RR>1	Risk in exposed greater than risk in unexposed (positive association)
RR<1	Risk in exposed less than risk in unexposed (negative association)

1.5.

Three conditions should be met:

1. When the cases studied are representative of all people with the disease in the population from which the cases were drawn;

2. When the controls studied are representative of all people without the disease in the population from which the cases were drawn;

3. When the disease being studied is not a frequent one.

1.6.

	Disease + (number of persons)	Disease − (number of persons)
Exposed	j	k
Not exposed	m	n

Write the appropriate formula for each measure:

    A.   Incidence proportion among exposed = j/(j+k)

    B.   Incidence proportion among not exposed = m/(m+n)

    C.   Relative risk =(j/(j+k))/(m/(m+n)

    D.   The exposure odds of the cases = j/m

    E.   The disease odds of the not exposed = m/n

1.7. Hypothetical cohort study:

	Myocardial infarction (MI) develops	MI does not develop	Total	Incidence per 1000 persons
Smoke cigarettes	90	2910	3000	30
Do not smoke cigarettes	80	4920	5000	16

A.   In population of smokers how much of MI that they experience is due to smoking?

30 – 16 = 14 cases per 1000 of the 30 per 1000 in smokers.

B.   How much of the MI could be prevented if they did not smoke? 14 cases per 1000 of the 30 per 1000 of MI that smokers experience.

C.   What be the total impact of a prevention program on the community (total population)? Incidence in total population = 170/8000=0.0213 x 1000 = 21,3 per 1000 population. Proportion of cases in the population attributable to the exposure = (21.3–16.0)/21.3=0.249 x 100 = 24.9%

CHAPTER 2. Answers

2.1. Sampling error is the part of the estimation error of a parameter from a sample caused by the random nature of the sample. Random error can be reduced by increasing the sample size. Selection bias is a systematic error that causes a distortion in the estimate of a parameter and is caused by the manner in which the subjects are selected from the total population into the sample. Bias cannot be reduced by increasing the sample size.

2.2. *Selection bias* is a distortion in the estimate of association between risk factor and disease that results from how the subjects are selected for the study. Selection bias could occur because the sampling frame is sufficiently different from the target population, or it could occur because the sampling procedure cannot be expected to deliver a sample that is a mirror image of the sampling frame.

Information bias occurs when the information obtained from study subjects is systematically inaccurate regarding the disease or exposure under study.

2.3. In a study of whether factor X is a cause of disease Y, we say that a third factor, factor C is a confounder if: factor C is known risk factor for disease Y, factor C is associated with factor X, but is not a result of factor X.

CHAPTER 3. Answers

3.1

    A. Experimental

    B. Observational

    C. Experimental

    D. Experimental

    E. Experimental

    F. Observational

3.2

    A.   Prospective cohort

    B.   Case-control

    C.   Prospective cohort

    D.   Cross sectional

    E.   Case series

    F.   Clinical trial

3.3

    A.   Therapeutic trial

    B.   Preventive trial

CHAPTER 4. Answers

4.1.
- A.   Prevalence;
- B.   Both;
- C.   Prevalence
- D.   Incidence

4.2.

Age	Annual Incidence	Prevalence	Duration=Prevalence/Annual Incidence
0–5	6/1,000	29/1,000	4.8 years
6–16	3/1,000	32/1,000	10.7 years
17–44	2/1,000	26/1,000	13.0 years
45–64	1/1,000	33/1,000	33.0 years
65+	0	36/1,000	33.0 years
Total	3/1,000	30/1,000	10.0 years

A.   The table shows the approximate annual incidence and prevalence rates for asthma. Incidence falls with increasing age, illustrating the fact that there are more incident cases in childhood than in adult. But the prevalence stays fairly stable over the entire age span, indicating that asthma tends to be chronic and is especially chronic among older individuals. In other words, with no or few incident cases the prevalence is as high as 30 cases per one thousand inhabitants

4.3.

A. $\mathrm{OR} = \dfrac{1{,}556 \times 14{,}532}{14{,}974 \times 694} = 2.18$

B. This odds ratio suggests that smokers are more likely than non-smokers to have low birth weight babies.

CHAPTER 5. Answers

5.1.

      A.   b;

      B.   a;

      C.   c;

      D.   b;

5.2.

    A.   **a;**

    B.   **c;**

    C.   **c;**

    D.   **b**;

    E.   **c**;

    F.   **b**;

    G.   **c**;

    H.   **c**.

5.3.

OR = 0.77, meaning that oral contraceptives do not increase the risk of ovarian cancer since the result is close to 1.0. Confidence intervals are needed to be more exact.

CHAPTER 6. Answers

6.1.

    A. b;   B. a;   C. b;   D. c.

6.2.

A. F;   B. F;   C. T;   D. T;   E. T;   F. F;   G. T

6.3.

The statement means that a 45-year-old male free of prostate cancer has a probability of 0.05 of developing prostate cancer over the next 10 years if he does not die from any cause during the follow up period.

6.4.

Smaller, because the 5-year risk involves a shorter time period for the same person to develop prostate cancer.

6.5.

    A.   FORWARDS

    B.   RETROSPECTIVE

    C. $27/75 = 0.36$

    D. $14/81 = 0.17$

    E. $0.36/0.17 = 2.1$
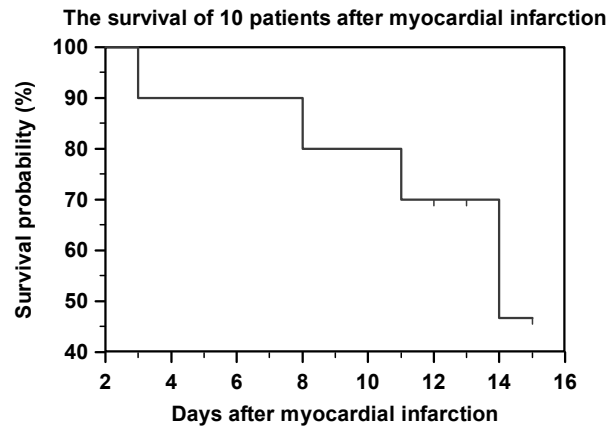
6.6.

A.3;

B.1;

C.0.5

6.7.

    A.   Prevalence 100/1000 = 0.1

    B.   3 year incidence 15/900 = 0.017

    C.   Annual incidence/risk 0.017/3 = 0.006

    D.   The lung cancer specific mortality risk for this cohort is 50/1000 = 0.05

    E.   The all-cause mortality risk is 200/1000 or 0.2

    F.   The case fatality risk for 100 lung cancer patients in the initial cohort is 40/100 or 0.4

6.8.

A. Survival time in days after myocardial infarction in ascending order	B. Number still in the study at start of the day	C. Number of death	D. Number censored	E. Proportion surviving until end of interval	F. Cumulative proportion surviving
3	10	1	0	$1 - 1/10$	0.90
8	9	1	0	$1 - 1/9$	0.80
11	8	1	0	$1 - 1/8$	0.70
12	7	0	1	1	0.7
12	6	0	1	1	0.7
12	5	0	1	1	0.7
13	4	0	1	1	0.7
14	3	1	0	$1 - 1/3$	0.47
15	2	0	1	1	0.47
15	1	0	1	1	0.47

The survival of 10 patients after myocardial infarction

CHAPTER 7. Answers

7.1. For a disease to be suitable for screening it must be serious, treatment given before the symptoms develop must be more efficacious in terms of reducing morbidity and mortality than that given after the development of clinical manifestations of the disease, and the prevalence of preclinical disease must be high among the screened population.

7.2. A good screening test should be highly sensitive (to produce few false negative results), highly specific (to produce few false positive results), simple, safe, inexpensive, acceptable to patients and clinicians, both

7.3.

A. T

B. F

C. F

D. T

7.4.

Screening test	Pancreatic cancer (confirmed by MRI and biopsy)		Total
	Cancer confirmed N=177	Cancer not confirmed N=64 633	
Positive	200	600	800
Negative	50	9,150	9,150
Total	250	9,750	9,750

$$\text{Sensitivity} = \frac{200}{200 + 50} = 0.8 \ \text{ or } 80\%$$

$$\text{Specificity} = \frac{9,150}{9,750} = 0.94 \ \text{ or } 94\%$$

7.5.

    A.   0.4;

    B.   2.36;

    C.   0.44

7.6.

ILBP	AS	
	Present	Absent
Present	95	10
Absent	5	90
	100	100

A. Prevalence = $\dfrac{100}{1000} = 0.1$

B. Sensitivity = $\dfrac{80}{100} = 0.8$

C. Specificity = $\dfrac{950}{1000} = 0.95$

D. PV(+)= $\dfrac{a}{a+b} = \dfrac{80}{80+50} = 0.62$

E. PV(−) = $\dfrac{d}{c+d} = \dfrac{950}{970} = 0.98$

7.7.

    A.   Sensitivity = 95/100 = 0.95

    B.   Specificity = 90/100 = 0.90

CHAPTER 8. Answers.

8.1. Scientists generally base scientific hypotheses on previous observations that cannot satisfactorily be explained with the available scientific theories. Even though the words "hypothesis" and "theory" are often used synonymously, a *scientific hypothesis* is not the same as a *scientific theory*. For a hypothesis to be put forward as a scientific hypothesis, the scientific method requires that one can test it. A scientific theory is a set of principles that explain and predict phenomena. Scientists create scientific theories with the scientific method, when they are originally proposed as hypotheses and tested for accuracy through observations and experiments. Once a hypothesis is verified, it becomes a theory.

8.2. To demonstrate the feasibility of doing the study as proposed, to demonstrate that the investigator(s) have the ability and skills to conduct the proposed study and are aware of all limitations in the design, crystallize the project to the researchers themselves, give referees the possibility to review the project (especially for funding), inform and educate all those taking part in the project, ensure the main researchers do not forget any details of the plan in the course of the study, document the procedures of the project for the future.

8.3. Generally a hypothesis is used to make predictions that can be tested by observing the outcome of an experiment. If the outcome is inconsistent with the hypothesis, then the hypothesis is rejected. If the outcome is consistent with the hypothesis, the experiment is said to support the hypothesis. Researchers recognize that alternative hypotheses may also be consistent with the observations. In this sense, a hypothesis can never be proven, but rather only supported by surviving rounds of scientific testing and, eventually, becoming widely thought of as true (or better, predictive), but this is not the same as it having been proven. A useful hypothesis allows prediction and within the accuracy of observation of the time, the prediction will be verified. As the accuracy of observation improves with time, the hypothesis may no longer provide an accurate prediction. In this case a new hypothesis will arise to challenge the old, and to the extent that the new hypothesis makes more accurate predictions than the old, the new will supplant it.

8.4. An epidemiologic hypothesis is a testable statement of a putative relationship between an exposure and disease. The hypothesis should be clear, testable or resolvable, state the relationship between exposure and disease, limited in scope, not inconsistent with known facts, supported by literature, theory, references. The objective presented in the abstract 1 can pretend to the epidemiological hypothesis

CHAPTER 9. Answers

9.1

The mean of the ten numbers: 1, 1, 1, 2, 2, 3, 5, 8, 12, 17 is 55/10 = 5.5

Seven of the ten numbers are less than the mean, with only three of the ten numbers greater than the mean.

A better measure of the center for this distribution would be the median, which in this case is (2+3)/2 = 2.5. Five of the numbers are less than 2.5, and five are greater.

Notice that in this example, the mean is greater than the median. This is common for a distribution that is skewed to the right (that is, bunched up toward the left and with a "tail" stretching toward the right).

X

9.2

No – since the sample sizes are different, the p-values are not comparable, even though everything else is the same. (In fact, a larger sample size typically results in a smaller p-value)

9.3

There are two correct answers:

A.   There is a 3% chance of observing a difference as large as you observed even if the two population means are identical (the null hypothesis is true).

B.   Random sampling from identical populations would lead to a difference smaller than you observed in 97% of experiments, and larger than you observed in 3% of experiments.

9.4.

A. The p value is the probability of the data, assuming the null hypothesis is correct

B. $p < 0,05$ is an arbitrary convention that has taken on unwise, indiscriminate use

C. False rejections may occur at any level

D. P values cannot be used to indicate evidentiary support for a hypothesis without making certain assumptions.

E. Statistical significance is a phrase that has come to mean the null statistical hypothesis has been rejected at a given "significance level".

CHAPTER 10. Answers

10.1. The scatter plot between the two variables showing the relationship between them.
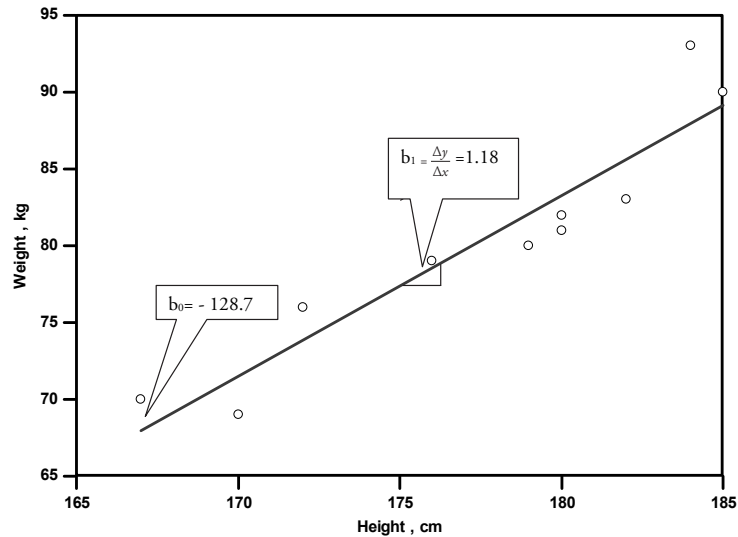
10.2. Both of them are normally distributed around the mean (Kolmogorov-Smirnov test for both variables are> 0.05).

10.3. There is a linear relationship between the two variables because the residuals are distributed normally (Kolmogorov-Smirnov test, p=0.83). We use a suitable computer program to calculate the residuals and to test their normality; usually it is a regression function. The statistical model for this is called – the method of ordinary least squares (OLS).

10.4. Use Excel spreadsheet application that comes as a part of Microsoft Office suite. Always plot independent variable on X axis. This way the height should be plotted on the X axis. The regression equation calculated for the two variables is: $y = b_0 + b_1 x$ or $y = -128.7 + 1.18x$.

10.5. There are two important coefficients in this equation: b_0 and b_1, what in this case are -128.7 and 1.18, in respective.

b_0 is a constant coefficient or intercept – it's where the line cuts Y axis. In our case it is -128.7. b_1 is a slope coefficient (1.18 in our case). The value of b_1 is the amount by which the weight of the person will change if the height of the person will change for one unit. The value of slope coefficient is calculated as follows: $= \dfrac{\Delta y}{\Delta x} = 1.18$. So if the height will increase for one unit the weight will increase app. for the same value -1.18. The slope should be statistically significant to generalize for population.

The linear regression model estimates how big proportion of observation can be described by this model. The bigger the better. The coefficient of determination for this model (R^2) is 0.88, meaning that height explains 88 % of the observed variation in weight.

CHAPTER 11. Answers

11.1. Explicit methods limit bias in identifying and rejecting studies. Conclusions are more reliable and accurate because of methods used. Large amounts of information can be assimilated quickly by healthcare providers, researchers, and policymakers. Delay between research discoveries and implementation of effective diagnostic and therapeutic strategies may be reduced. Results of different studies can be formally compared to establish generalizability of findings and consistency (lack of heterogeneity) of results. Reasons for heterogeneity (inconsistency in results across studies) can be identified and new hypotheses generated about particular subgroups. Quantitative systematic reviews (meta-analyses) increase the precision of the overall result.

11.2. It needs to be an analysis of evidence and not just be a literary review talking of selected studies.

11.3. For busy healthcare providers and decision makers, systematic reviews are important as they summarize the overwhelming amount of research-based healthcare information that is available to be read and synthesized. They also overcome some of the bias associated with small single trials where results may not be robust against chance variation if the effects being investigated are small. Finally, systematic reviews may overcome the lack of generalizability inherent in studies conducted in one particular type of population by including many trials conducted in varying populations.

CHAPTER 12. Answers

Exercise 1

X

A. Quantitative. In order for the effectiveness of the two drugs to be compared it would need to be measured.

B. Qualitative. The study aims to explore the role of the practice manager and will be describe a phenomenon. The fact that the study is conducted in only four practices also suggests an in depth study.

C. A descriptive study of experience suggests a qualitative approach. Also, the focus is boys with eating disorders and difficulty in locating a sizeable sample may be anticipated.

D. A national survey suggests a large scale study. The data could be collected using a questionnaire.

Exercise 2   ✗

    A.   Phenomenology – the study seeks to explore and describe a phenomena.

    B.   Ethnography – to inform the development of a service for a particular cultural group, the research would seek to understand the beliefs and practices of the culture.

    C.   Grounded theory – if we can understand and describe what advocacy actually means in primary health care, the new knowledge can be incorporated into practices and policy.

    D.   Case study – the polyclinic is a "case", a unit of study.

Exercise 3   **X**

1. At this stage, all the information is new and everything is potentially interesting.

2. You are likely to have identified some or all of the following categories: – feelings of the respondent – fear, embarrassment or shame, surprise – beliefs about people with mental illness – who they are and how they appear – expectations of the hospital – attitude towards smoking – concerns about security

3. Your categorisation may be broader or narrower than this; consequently the numbers of categories may be different. But isn't it interesting that so many categories can be generated from one page of transcript?

4. As an example, expectations and beliefs might be one or two categories.

5. For example, feelings of the respondent might be a major category and the different feelings could be minor categories. You would decide as further interview transcripts were analysed.