An understanding of the basic elements of research is essential for good research practices. Among the most important elements to be considered are variables, associations, sampling, random selection, random assignment, and blinding.

**1) THE RESEARCH QUESTION:**

The best research question should specify just one measurable result, as well as all the conditions and important variables.

The question contains population, manoeuvre or the conditions affecting the study population, and the results.

**2) HYPOTHESIS:**

Assumption, based on observations or reflections, which may lead to refutable predictions.

Also a conjecture drawn up so that it can be tested and refuted.

* **a) Alternative hypothesis:**

Generally this is defined as stating that the null hypothesis (there are no differences) is not true.

If the objective of the trial is to compare a drug with the placebo, the null hypothesis will state that there is no difference between the two groups, and the alternative hypothesis that there is a difference.

* **b) Null hypothesis:**

Hypothesis that there is no difference between the two groups (for example, two treatments).

When both groups differ in substantially, the null hypothesis is very improbable.

**3) RANDOMIZATION:**

Ideally, a process which ensures that each member of a population has an equal chance of being included in the study sample. This is not always possible.

More importantly, randomization means that the patients of the study are allocated to treatment or placebo groups for the experiment, without taking into account any of the the patients’ characteristics or the desires of the study personnel.

**4) POPULATION:**

Every person who meets the inclusion criteria for the study.

Group of people with a defined characteristic (white females) or who meet a defined requirement (people who live in Andalucía).

The denominator in the calculation of a rate.

**5) SAMPLE:**

The people who met the inclusion criteria of the study and who actually are incorporated into the study. A sub-set of the population.

Selected fraction of a certain population, ideally representative of the reference population and of a large enough size. Subgroups of observations of the study population.

**6) SAMPLING:**

Selection of the subjects of a population who will take part in a study.

Once the population for a clinical trial has been defined, the participants should ideally be selected randomly, that is, in such a way that each potential participant of the population under study has the same probabilities of being included in the sample.

This rarely happens, because the people identified as elements of the reference population may have some characteristic which makes them specifically identifiable.

**7) MANOEUVRE:**

Any exposure or treatment which acts on the patients to produce a result.

**8) RESULT OR OUTCOME:**

The effect of a manoeuvre.

This term is also used to designate pre-determined variables in a clinical trial (for example, cardiovascular mortality + reinfarction, in clinical trials with patients who have suffered a myocardio infarction).

**9) INTERNAL VALIDITY:**

Are the results of a study valid for the patient population studied?

Degree to which the experimental and reference groups have been formed and compared in such a way that the differences observed between them in the dependent variables studied can be attributed solely to the intervention under investigation.

**10) EXTERNAL VALIDITY:**

Are the results valid outside the population of patients studied? Are the results of the studies done in men valid for women?

Degree to which the conclusions obtained with the sample of the population which participates in a study can be generalised to its reference population or to other populations, places, times and researchers.

**11) VARIABLE:**

Any attribute, phenomenon or fact which may have different values. A property of an individual which can be observed; for example, height, weight, sex; this property varies from one individual to another.

There are three levels of precision in the measurement of a variable: nominal, ordinal and interval.

The simplest is the nominal one: the values assumed by a variable at this level indicate simply categories (for example, sex, treatment group, etc).

The variables can also be rank ordered (ordinal grouping), for example, according to the level of dose, categories of treatment duration or according to social class.

The variables which can not only be ordered but which also permit the measurement of the distance between categories are called intervals (for example, height, weight, blood pressure, number of deaths).

The values which any variable can take, whether explanatory or response values, constitute the levels of the variables.

The nature of these levels is what determines the type of the variables.

Therefore the following types of variables exist:

* **Dichotomous:** a variable which can only take two possible values.
* **Nominal:** a variable which can take several values.
* **Ordinal:** a variable which can take several values and in which a ranking between the categories can be established.
* **Quantitative:** a variable which can take a numerical range of values.
* **Censored:** a quantitative variable which can take partial or incomplete information (time until something occurs in Analysis of Survival).
* **Binary variable:** variable or result which can only take one of two possible values.
* **Categorical variable** (synonym: discrete variable): Said of the variable which presents interruptions.
* A variable is categorical if, among several potentially observable values, there is a value which cannot be observed.
* Binary variables are a type of categorical variable, which can only take two values.
* **Continuous variable:** said of the variable which does not present interruptions; a variable is continuous if, between two observable values, there is always the possibility of there being another observable value.

Examples; age, weight, height, blood pressure.

* **Dependent variable:** variable whose value depends on the effect of another or other variables (independent variables).
* Manifestation or result, the value of which is explained or justified by the influence of independent variables, especially the regression model.
* In statistics, the variable which can be predicted by means of a regression equation.
* Independent variable: characteristic whose influence on a fact or a manifestation is assumed (the dependent variable).
* In statistics, the independent variable is one of the elements of a regression equation.
* **Intermediate variable** (synonym: contingency): variable present in a causal chain from an independent variable to another dependent one.

It determines the dependent variable, but at the same time is the object of modification by the independent variable and simultaneously associates with both.

**12) RANDOM ALLOCATION:**

In a study in which two or more types of treatment are compared, it is advisable for the groups formed to be similar in all the prognostic characteristics except in the treatment received, so that any difference in the clinical course which is recorded can be attributed to the different treatments administered (and only to them).

Random allocation consists of distributing each participant to one of the treatment groups by means of a random method, so that each subject has exactly the same probabilities of forming part of one treatment group or another.

The patients of one group have, on average, the same probability of possessing a certain characteristic as those of the other group; this occurs with all the prognostic factors, known or unknown.

When the number of patients included in a clinical trial is limited, random allocation may determine the formation of somewhat different groups; this is less probable the greater the number of patients.

With a view to evaluating whether random allocation has been carried out by chance, the distribution of the prognostic characteristics known in each group should be compared at the end of the trial.

Most publications about clinical trials contain a comparative table of these characteristics.

In order to prevent unequal distribution of the prognostic characteristics in each group, an allocation in block can be carried out or the known differences can be corrected with a stratified analysis or with a regression analysis.

Random allocation is the defining characteristic of controlled clinical trials, and differentiates these trials from cohort studies.

* **Stratified random allocation:**
* Method whereby, before random allocation, the patients are included in subgroups (strata) of similar prognostic characteristics and are then allocated randomly, separately for each stratum, to each of the types of treatment.
* In this way, the final treatment groups and subgroups formed are comparable, at least in terms of the factors which have been considered when forming the blocks or strata.
* For example, the risk of cardiovascular complications associated with high blood pressure is greater in diabetic patients, but it might also happen that a group of antihypertensive drugs may produce more undesirable effects among the diabetic patients.
* In a clinical trial on antihypertensive drugs, a general random allocation can be made, or diabetic and non-diabetic patients may be allocated separately to each type of treatment; in both cases, the results in diabetic and non-diabetic patients could be analyzed separately.
* Nevertheless, the results will have greater validity if the allocation has been carried out separately, than if only an analysis of subgroups were made.
* **Block random allocation**:

Random allocation applied to successive groups of patients (for example, groups of four, groups of ten, etc), with a view to preventing imbalances between groups in the case of intermediate analysis or interruption of the trial.

* **Random allocation by minimization: This consists of :**
1. Identifying the relevant variables which are to be distributed in a well-balanced way to both treatment groups (for example, in a clinical trial on the treatment of acute myocardial infarction, age, smoking habit, antecedents of coronary heart disease and diabetes).
2. Forming strata according to the variables of interest (in the example, patients under 70 or 70 and over, smokers or non-smokers, with or without antecedents of coronary heart disease, and diabetics or non-diabetics).
3. Counting, for each stratum, how many patients have been allocated to each of the types of treatment.
4. Adding the patients included in each of the stratum in one or other type of treatment.
5. Allocating new patients to the type of treatment in which the subjects of their characteristics are least represented.
* **Alternate allocation:**
* In a clinical trial with two groups, method of treatment allocation in which the first patient receives treatment A, the second patient receives the alternative (B), the third patient treatment A, the fourth one B, and like this successively, in a predictable way.
* This type of allocation is not optimum (it generally means that the researcher knows which technique, A or B, corresponds to the next patient) and therefore can give rise to unequal groups and it is rarely compatible with maintaining the double-blind character of a trial.
* **Concealed allocation:**

Allocation process described in such a way that it gives reason to believe that adequate measures were taken to conceal the allocation of patients to each group from the people in charge of evaluating the patients (for example, centralized random allocation, use of numbered, opaque and sealed envelopes contained in a sealed bag, use of numbered or encoded flasks and other recipients or other descriptions with elements which ensure concealment).

**13) INFORMED CONSENT:**

Formal procedure to apply the principle of autonomy.

It should include at least three elements: voluntariness, information and comprehension.

Voluntariness means that the subjects can decide freely if they want to participate in a study, without there being persuasion (that is, when they are induced to “freely accept” attitudes, values, intentions or actions), manipulation, real options or their perception of choice, or coercion.

The voluntary character of consent can be infringed when it is requested by people in authority or who have great influence over the subject of the research and when there is not “enough” time for the subject to reflect, consult and decide.

The information should be understandable and should include the objective of the study and its procedure, the potential benefits and risks, the possible inconveniences derived from the participation and the possibility of withdrawing from the study at any time without causing prejudice.

In this context, understanding of the information is basic, depending not only on the information itself, but also on the competence of the person who receives it, that is, their capacity to understand.

**14) REPRODUCIBILITY:**

With reference to a test, degree to which the results are identical or closely similar every time the test is carried out.

**15) ABSOLUTE RISK:**

Term used with different meanings depending on the school of epidemiology, which should consequently be avoided.

**16) ATTRIBUTABLE RISK:**

Synonym: etiologic fraction.

Proportion of all the cases of a disease which are attributable to a certain risk factor.

**17) ATTRIBUTABLE RISK AMONG EXPOSED CASES:**

Proportion of exposed cases which is attributable to a certain risk factor.

**18) RELATIVE RISK:**

Ratio between the risk of suffering from a certain disease among the individuals exposed to a certain risk factor and the risk of suffering from it among those not exposed.

If a is the number of exposed individuals with the disease, b the number of exposed individuals without the disease, c the number of unexposed individuals with the disease and d the number of unexposed individuals without the disease, the risk among the exposed individuals is a/a + b, the risk among the unexposed individuals is c/ c + d and the relative risk