

# How to plan an experiment

## I. Randomization: current fad or (ever)lasting fashion?

---

Cezary Watała

Department of Haemostatic Disorders, Medical University of Lodz,  
Medical University Hospital No. 2, Lodz, Poland

**Submitted:** 6 January 2006

**Accepted:** 10 January 2006

Arch Med Sci 2006; 2, 1: 58-65

Copyright © 2006 Termedia & Banach

**Corresponding author:**

Prof. Cezary Watała  
Department of Haemostatic  
Disorders  
Medical University of Lodz  
Medical University Hospital No. 2  
Zeromskiego 113  
90-549 Lodz, Poland  
Phone: +48 42 639 34 71  
Fax: +48 42 678 75 67  
E-mail: cwatala@csk.umed.lodz.pl

### Abstract

All those who like experimentation are sooner or later faced with a need of random selection of elements or objects of interest that they want to study (persons, patients, animals, cells, etc.). Randomization, a basic requirement in appropriate planning of experiment, may be performed either to select series of randomly assigned elements/objects or to allocate the studied objects to a given group, medical or diagnostic procedure, treatment protocol, etc. The principal advantage of random selection is to minimize effects of bias and confounding variables, two fundamental threats known to weaken research credibility. Simple, unrestricted random selection can lead to undesirable imbalance in baseline characteristics, thus affecting any credible reasoning. Restricted randomization (with blocking or stratification) includes procedures used along with random sampling that help to achieve balance between study groups in their baseline characteristics or in size. For all researchers conducting biomedical studies randomization ensures a straightforward and reliable analysis of the outcomes, and enables any further generalization of their findings.

**Key words:** experiment, randomization, (bio)medical study, bias, confounder, blocking, stratification.

---

One of the researcher's fundamental natural behaviors and expectations is to extrapolate the outcomes and conclusions of a certain study performed on selected examined groups of objects to larger, more general population. The easiest solution, we might think about *ad hoc*, would be to include under study as many objects as possible, to make our generalization possibly the most reliable. For numerous reasons, including the economic reasonability, such an approach is certainly beyond any limits we could rationally cope with. Any other solutions force us to approximate, round up and rely on certain assumptions.

**Randomization** is one of the most common basic assumptions used to enable any further generalization of our findings. Any conclusions derived from studies devoid of appropriate randomization protocol may refer exclusively to the group of elements under study and cannot be any more universal or extrapolated towards larger groups. In other cases, our reasoning and conclusions may very likely be false.

Below, you will find arguments that the demand of randomization in research is not a pure personal whim or addiction. It is a need for mature scientific research. This need has been glorified through years by unambiguous scientific evidence, showing that our research is as good as the imagined

experimental design of what we are planning to perform. Requirement of randomization, previously more a perplexing demand, recently evolved to the status of ally for researchers conducting biomedical studies. Thus, this fundamental step in experimental design truly became a fact and is no longer regarded as merely the fashion, represented by a minority of researchers, who do not dislike too much a statistical approach to experimentation.

### Randomization – what do we need it for?

The procedure of randomization means the selection of elements or objects under study (persons, patients, animals, cells, etc.) by pure chance, without a possibility to predict a given choice. Random selection is a basic requirement in appropriate planning of experiment and it ensures a straightforward and reliable analysis of the outcomes. It means for the researcher an objective and reliable selection. To give a brief answer to the question “why do we need to randomize?” we may state that we insist on selecting the objects under study at random in order to express general conclusions (referring to a general population) based on particular outcomes (referring to a small fragment of a general population). By random sampling we guarantee that the observed characteristics closely reflect the characteristics of a whole population; we say that our group is statistically representative to a general population. Thus, we save our energy, time and – very importantly – money, when investigating a “well-cut,” representative portion instead of a whole population [1].

Randomization may be performed either to select series of randomly assigned elements/objects or to allocate the studied objects to a given group, medical or diagnostic procedure, treatment protocol, etc.

Below are typical examples of random selection(s) in medical studies:

1. 500 women over 60 years of age, with family names beginning with G through P, inhabiting Poznań City, were randomly selected from the alphabetical lists delivered by Poznań registration offices. Each individual was sent a questionnaire, and the study was based on the assumption of a reply from at least 300 responders.
2. Fifty consecutive patients with pulmonary edema were randomly allocated to the group receiving i.v. infusion of nitroglycerine (NG) or to the group receiving oral NG administration repeated every 5 min. Random selection to either group was performed by drawing balls from the bag containing initially 25 white and 25 black balls.
3. In a multi-center trial the body height was recorded in active working adults applying to regional outpatient clinics of occupational medicine. In each clinic the first 100 incoming patients were enrolled. Since 99 clinics participated in the study in a whole country, data on 9900 patients were recorded.

### Tools for random selection

The random selection may be secured by several techniques, including:

- tossing a coin or rolling a die;  
for those interested in rolling more than one die at a time or using more than 6-sided dice, we recommend two web addresses:

– <http://images.google.pl/imgres?imgurl=www.neophyre.com/msc/rdr1.jpg&imgrefurl=http://www.neophyre.com/msc/rdr3.html&h=249&w=177&sz=14&tbid=y4yJ5pUQJr8j:&tbnh=105&tbnw=75&start=1&prev=/images%3Fq%3Drandom%2Bdice%26hl%3Dpl%26lr%3D%26ie%3DUTF-8%26oe%3DUTF-8%26sa%3DG>

– [http://images.google.pl/imgres?imgurl=faculty.haas.berkeley.edu/mss/tools/dice\\_1a.gif&imgrefurl=http://faculty.haas.berkeley.edu/mss/tools/tools\\_ranndice.htm&h=302&w=304&sz=3&tbid=iA1K\\_CA V7ckj:&tbnh=111&tbnw=111&start=13&prev=/images%3Fq%3Drandom%2Bdice%26hl%3Dpl%26lr%3D%26ie%3DUTF-8%26oe%3DUTF-8%26sa%3DG](http://images.google.pl/imgres?imgurl=faculty.haas.berkeley.edu/mss/tools/dice_1a.gif&imgrefurl=http://faculty.haas.berkeley.edu/mss/tools/tools_ranndice.htm&h=302&w=304&sz=3&tbid=iA1K_CA V7ckj:&tbnh=111&tbnw=111&start=13&prev=/images%3Fq%3Drandom%2Bdice%26hl%3Dpl%26lr%3D%26ie%3DUTF-8%26oe%3DUTF-8%26sa%3DG)

- pulling out different colored balls from a bag; each color representing a given group of allocation (e.g. treatment); the balls taken out may then be given back or not to a bag; in the first case we ensure a continuous selection at random; however, we cannot assure the equal size of groups; on the other – when placing the equal numbers of balls representing each of the groups – we gradually lose randomness – with each subsequent pulling we make the total number of possible selections smaller, thus reducing randomness;
- selection of envelopes containing an indication of a medicine or placebo;
- computer software;

It is probably the most favored approach nowadays, although not always the most suitable for our practical needs. Various computer statistical packages have been developed to look for non-randomness and to offer well researched algorithms (“random” number generators) providing extremely long series of numbers for which there is an infinitesimally small probability of finding a repeating pattern. Importantly, most “quick” random number generators, including those supplied with computer language compilers, often use over simple methods, which produce sequences of numbers with repeating patterns; these are unacceptable for statistical use and should be avoided.

Random number generators require the so-called seed number [the number used to start with when generating (pseudo-)random series of numbers]. This specific “anchor” number is provided by a computer or given by the user. However, even when you have the opportunity to enter your own seed number, you

should rather use the default one (given by the program) in most cases. Why? We should note that each seed generates its own series of numbers and that series is the same if you use the same seed again. In other words, if the generator is given the same seed each time it is called then it will produce the same series of numbers. This is not acceptable for many purposes, and therefore, the random number generator needs to be reseeded each time it is used, to lower the risk of using the same (pseudo-)random number series for different randomizations procedures. How is such a risk minimized by computer software? The algorithm of (pseudo-)random number generation is based upon our computer's clock. Very often, the computer-selected seed is the number of hundredths of a second which have elapsed since the last midnight. Hence, it seems highly improbable that any software will produce the same "random" sequence more than once [2, 3];

- a simple "wild guess".

In the case of very large studied groups the random selection may be replaced by a simple inclusion of consecutive patients. How to make sure – at any given stage of our random selection – that we are really close to be "incidental" at such a selection? Probably, the easiest practical way is to verify the distribution of some continuous variables assigned to selected objects. A normal distribution of data may be considered a very good recognition that our objects are randomly selected [4, 5].

### Advantages of randomization

We probably profit most from using randomization by minimizing two fundamental threats that are known to weaken the credibility of our research:

- bias, and
- confounding variables.

The **bias** is a kind of a systematic error leading to an incorrect estimate (underestimate or overestimate) of the investigated effect or association. Numerous factors (of researcher's interest or not) can bias the outcomes of our study in such a way that they cancel out, reduce or amplify a real effect(s), which we are trying to describe.

There are several types of bias, and the most commonly encountered are:

- selection bias; occurs e.g. when investigating inhabitants of a small alpine village to describe the characteristics of central European population, or when we try to allocate the most healthy individuals to the treatment that we intend to prove is the best;
- observation bias; may occur when collecting data (e.g. it is more likely on questioning that healthy subjects underreport their alcohol intake compared to patients with coronary heart disease; women with complicated pregnancies are more likely to

declare the prior use of oral contraceptives), interviewing participants (e.g. the style of interview may provoke some answers over others, different responders may give different answers to the same questions) or failure to classify properly;

- systematic bias; occurs e.g. when performing the tests on treatment 1 in winter, while the tests on treatment 2 in summer;
- accidental bias; occurs e.g. when taking first out of the cage(s) the laboratory animals given one treatment, and the remaining animals for another treatment;
- inability to follow up; occurs e.g. when patients in poorer clinical condition are not able to continue their participation in a study, whereas healthier individuals are more likely to complete the protocol;
- cheating by the experimenter; may not always be badly intentioned by a researcher; occurs for example in the cases when the experimenter decides to (a) simplify his/her task and make life a little easier by performing a test first in all participants with families beginning with C, and then in all with families beginning with D; (b) select a patient to a trial if the participant would particularly benefit from a tested medical treatment; (c) give an extra dose of acetylsalicylic acid to those patients who are at the highest risk of re-occlusion; (d) balance the numbers of selected objects over some nuisance variable(s) without troubling a statistical expert.

We can of course minimize or combat the unwanted bias through:

- selecting more than one reference (control) group;
- standardizing of our observations using blinding/masking procedures for subject/observer (single blind), both subject and observer (double blind) or subject, observer and analyst (triple blind);
- using multiple sources of information and verifying their corroboration;
- using dummy variables with recognized interactions and associations with other variables (parameters).

**Confounding variables** (confounders) are those variables that are associated with both the cause (modifying factor, exposure or risk factor in epidemiology) and the effect(s) (outcome, result or consequence, disease in epidemiology). Although confounders often predict the effect(s) (disease) very well and they may (or may not) be a part of the real association between a cause and an effect (exposure and disease), they are not a real part of what we are after when investigating a given association of interest. For example, when examining the association between sedentary behavior or occupation and development of cardiovascular complications, we have to control in our multiple regression analysis model for both obesity and smoking, since these two confounders are unequally distributed between people who show

sedentary behavior. Obese people tend to spend more time in the sitting position, while smokers are usually more mobile and physically active. Similarly, light hair color well predicts coronary heart disease when analyzed by multivariate statistical methods, because it is unequally distributed between people of different ages: elder people, who are at a higher risk of heart disease, usually have gray hair, just opposite of younger, who are at much lower risk for heart disease and are very seldom gray. Thus, light hair color confounds our thinking about heart disease because it is not a real cause of this disease. Other confounding variables and very typical of biomedical research are e.g. age structure or sex proportions of the studied population.

The optimal situation for each researcher would be of course to maximally control (or standardize) for the occurrence of confounders, which is possible only when they are known and measurable. The most common strategies to reduce the influence of confounding variables are:

- randomization: to ensure random distribution of confounders between the groups under study;
- matching and adjustment: to ensure equal distribution of confounders and group homogeneity; the appropriate adjustment may be distorted by choosing a standard;
- restriction of inclusion criteria: to ensure the exclusion of individuals with (numerous) confounding factors, each of which may constitute the risk bias in itself;
- stratification: to ensure that potential confounders are distributed evenly within each stratum (smaller part of the population under study; see below);
- *post hoc* multivariate analysis: to standardize (control) for those confounders that we are able to identify and measure; the aim of these analyses (so-called **adjusted analyses**) in biomedical research is to control (adjust) for some baseline imbalances in important patient characteristics. A similar approach may be used also to adjust significance values to take account of multiple testing known to increase the probability of making a type I error, i.e. attributing a difference to an intervention when a pure chance is the more likely explanation [1, 2].

### Types of randomization

Randomization may be either simple or restricted. Unrestricted, **simple randomization** can lead by chance to undesirable imbalance in baseline characteristics (the values of demographic, clinical, laboratory or other variables collected for each participant at the beginning of the study, but before the intervention is administered), including the so-called prognostic variables (those variables that are prognostic in the absence of intervention). This of course may affect the outcomes and weaken the credibility of researcher's reasoning. We try to protect our study against such imbalances by using either a restricted randomization model (e.g. stratification) or minimization.

**Restricted randomization** refers to any procedure(s) used along with random sampling that helps to achieve balance between study groups in their baseline characteristics or in size. To achieve these goals we have two statistical "tricks" at our disposal: blocking or stratification. Regardless of which of these two is used, we have to be aware that improved balance always comes at the cost of reducing the unpredictability of the sequence of random numbers. Thus, we always achieve one aim (reducing imbalance) at the cost of the other (reducing randomness) [1-3, 6].

### Blocking

This operation is aimed to ensure close balance in the size of each group at any time during the study. *Blocking* renders that comparison groups will be of approximately the same size. Blocking performed on every 20 participants would result in the allocation of 10 individuals to one treatment and 10 to another. Each block contains a random sequence of the interventions 1 and 2, however, it is possible to deduce some of the next treatment allocations if we know the block size. Therefore, although we achieve better within-group homogeneity (matching), we encounter the reduced randomness and unpredictability. To at least partly ameliorate this problem, we may blind the interventions, use blocks of larger sizes, or use blocks of randomly varying sizes.

### Stratification

Stratification is another approach to overcome the problem of imbalance concerning baseline characteristics of groups under study. This is a particular problem in small groups of participants, which by chance may not be well matched for some baseline parameters, like age, sex proportions, therapy, stage of disease, etc. We can try to avoid such imbalances by improving the matching of a distribution of a given variable among several compared groups; however, there is always a risk of how to do it without sacrificing the advantages of random sampling and losing study credibility. Stratified randomization is achieved by performing a separate random sampling within each of examined subsets of participants. It is a kind of restriction implemented in randomization procedure. The random assignment (allocation) of participants occurs within smaller groups defined by some basic parameters in a population. For example, we can stratify for age, sex, severity of disease or research center (like in multicenter trials): each parameter used to create the subset of data is considered a stratum. The appropriate stratification requires, of course, the prior blocking within strata (adjusting sample size), since without blocking the procedure is ineffective. Thus, the stratified random sampling ensures that not only the numbers of individuals

receiving each intervention are closely balanced within each stratum, but also that there is a good balance in participants' characteristics across intervention groups. Stratified blocked randomization enhances the power of small trials by reducing the variation in outcome due to chance disproportions in baseline parameters. It is of little benefit in large trials (1000 subjects and more), where chance assignments ensure nearly even distributions of baseline parameters. The limitation of stratified blocked randomization may be the relatively small number of baseline parameters (usually not more than 2-3) that can be balanced.

Below, are the most popular experimental demands for using randomized sampling (generators for some of them may be found at given web addresses) [6]:

1. Randomizing a simple series of data. It is a very commonly used procedure to select a smaller population from a larger one. For example, we can use this model in a medical research if we want to randomize the selection of a source population of 10 participants from a target population of 50 individuals. The useful tool for doing this can be found at <http://www.wcrl.ars.usda.gov/cec/java/r91-fig3.htm>:

24 7 30 29 44 18 28 32 38 36

2. Allocating samples to different treatments; may have several variants, including those presented below:

- random allocation to two independent groups; the simplest model used in randomized controlled trials (RCT) to allocate some subjects to receive the new treatment and the other subjects to receive the control treatment (e.g. reference drug or placebo);
- matching of samples in a sequential order evenly to several treatments; generator, which can be found at: <http://www.wcrl.ars.usda.gov/cec/java/sequence.htm>, takes a sequence of numbers representing a series of samples and divides them at random into several groups (each containing the same number of samples) representing different treatments. For example, to allocate 25 volunteers to 5 treatments with different inhibitors of HMG-CoA reductase (statins) we may use the following experimental design:
  - statin 1: 1, 6, 15, 18, 21,
  - statin 2: 4, 9, 11, 20, 25,
  - statin 3: 2, 8, 14, 16, 24,
  - statin 4: 5, 10, 12, 17, 22,
  - statin 5: 3, 7, 13, 19, 23.

- selecting items representing  $n$  categories to  $k$  treatments; generator can be found at: <http://www.wcrl.ars.usda.gov/cec/java/seq2.htm>. *Example:* we plan to investigate the effects of 5 different hypoglycemic agents in outpatient clinic individuals registered in local diabetological units

in 4 villages (A-D): 41 persons from the village A, 21 from the village B, 11 from the village C and 24 from the village D. Random selection may give a following solution:

*treatment 1:* A13 A39 A34 A30 A27 A11 A14 A32 B13 B9 B21 B8 C11 C7 D7 D17 D24 D18

*treatment 2:* A25 A4 A33 A35 A12 A3 A22 A36 B17 B15 B18 B5 C10 C1 D16 D21 D10 D22

*treatment 3:* A7 A21 A26 A41 A6 A28 A24 A18 B20 B19 B14 B1 C4 C9 D8 D9 D6 D14

*treatment 4:* A9 A23 A37 A5 A29 A10 A2 A31 B6 B16 B12 B10 C5 C2 D15 D1 D19 D13

*treatment 5:* A15 A20 A19 A1 A8 A40 A16 A38 B4 B2 B11 B3 C6 C8 D3 D11 D2 D4

- block randomization to  $k$  treatments; random allocation in blocks is made in order to keep the sizes of treatment groups similar; we have to remember to specify a sample size that is divisible by the chosen block size, and a block size that is divisible by the number of treatment groups (treatments).

*Example:* we want to allocate 20 cardiovascular patients to the treatments with two antiplatelet (AP) drugs: either acetylsalicylic acid or Clopidogrel (treatment 1 or 2); we decide to match up the participants in blocks of 4 patients each in order to limit the influence of confounding factors (e.g. sex, age) from hiding a real difference between two treatments.

block 1	
patient	AP drug
1	2
2	1
3	1
4	2

block 2	
patient	AP drug
5	1
6	2
7	1
8	2

block 3	
patient	AP drug
9	1
10	2
11	2
12	1

block 4	
patient	AP drug
13	2
14	1
15	2
16	1

block 5		AP drug
patient		
17		1
18		1
19		2
20		2

How to assort the suitable block size may be a real dilemma for the beginners with randomization. On the one hand, the advantage of small block sizes is that it is easy to achieve almost perfect balance, which means that treatment group sizes are very similar. However, we also remember the principal inconvenience of such an approach: with small block sizes we are likely to guess some allocations, so we lose the advantage of randomness and blinding. On the other hand, we may decide to use large block sizes; however, the compromise is equally difficult to be met. The only good alternative is to use blocks with sizes specified by random sequences.

Blocking compensates for situations where known factors (like sex or age) other than the status of a treatment group are likely to affect the studied association. Using this analytical approach we control for the fact that studied patients/subjects (our experimental units) are not homogeneous with regard to some factors (other than the status of treatment group), which are likely to affect the outcome of our study. Accordingly, our task is to (a) collect together homogeneous elements studied in our experiment (e.g. patients, laboratory animals) into a block, and then to (b) assign treatments at random within a block.

*Example:* we study the in vitro effect of different antiplatelet drugs on blood platelet reactivity evaluated with whole blood impedance aggregometry with collagen as an agonist; our outcome is the observed platelet aggregation of whole blood samples; blocks are individuals, who donated a blood sample; treatments are different antiplatelet drugs, by which portions of each of the blood samples are processed.

3. Randomizing the pairs of control and intervention enables a random allocation into the groups of intervention (with treatment) and control subjects (no treatment); used in crossover-designed studies, in which participants experience both intervention and control treatment at some stage during the study; the randomization in pairs allows to plan the order of treatment. The possibility of some carry-over effect(s) of intervention on control treatment may result in a bias in this type of study design.

*Example:* when randomizing intervention-control pairs in the study of the effects of acetylsalicylic acid therapy on serum salicylate levels in 10 hypercholesterolemic patients, we can use the following design:

1 intervention – control  
 2 control – intervention  
 3 intervention – control  
 4 intervention – control  
 5 control – intervention  
 6 control – intervention  
 7 intervention – control  
 8 control – intervention  
 9 intervention – control  
 10 control – intervention

4. Uniform random sampling of time periods or distances; generator can be found at: <http://www.wcrl.ars.usda.gov/cec/java/time.htm>. *Example:* we are to measure the flow rate of blood in a viscosimeter; we may do that by counting drops of blood collected in a beaker; for our convenience we choose to perform 5 repeats of counting, each lasting for 5 sec at the most, and we need resting breaks between each episode of careful observation (with counting !) lasting for minimum 10 sec; finally we can devote no more than 5 min for the game. Our possible approach may be:

1:25-1:30 hr

2:28-2:33 hr

2:52-2:57 hr

3:25-3:30 hr

4:16-4:21 hr

It shows we are supposed to waste at least 3 min for loaf-about; all we need is 100 sec.

### Minimization

When analyzing small groups minimization procedure is the only acceptable alternative to classical random sampling. Minimization is claimed to be superior over typical randomization under such conditions, and approaches using minimization are considered methodologically equivalent to randomized studies. Its particular advantage is that minimization ensures balance between intervention groups for several factors included in basic characteristics of participants under a study. It enables the researcher to make small groups closely similar with respect to some initially chosen parameters essentially at all stages of the study. Using this strategy the choice of the first participant is truly random, while for each subsequent element the allocation of treatment is identified in such a way that it minimizes the imbalance occurring between the examined groups [1, 2].

### How to (properly) use the tool of randomization? – a brief summary

We have to remember that a correct design of randomization is only one element of our research. We fully benefit from that provided other demands – including conceptual and methodological approach

– are worked out. Before we start with our study/experiment we need first to write down a protocol describing what we are supposed to do in our experiment. When preparing a protocol we need to answer at least the following questions [2, 3]:

- What is the purpose/idea of the experiment?  
The answer should not be something vague, like: "to investigate new disease" or "to characterize patients with type 2 diabetes mellitus given some oral glycaemic agent" or "to find out about the effects of ACE inhibitor on an organism", because even in a perfectly worthwhile experiment a statistician would be unable to help with data analysis. It is much better to state a specific question that can be answered either in a dichotomic (yes/no) or quantitative manner, for example:
  - to test the hypothesis that there is no difference between two different doses of thienopyridines in cardiovascular patients;
  - to estimate how much more efficient is simvastatin than atorvastatin in reducing LDL cholesterol;
  - to elaborate a regression model describing the kinetics of ASA enzymatic hydrolysis in blood plasma in the presence of salicylates.
- What are the treatments supposed to be tested?  
The treatment is what we are doing with the studied objects and how we influence the elements from the population under study (the so called experimental units, e.g. subjects, patients, animals, etc.). We need to give a precise description of the treatments intended to apply to the experimental units (e.g. 4 mg ASA per kg body weight, given peritoneally, every day). The description should contain complete technical details, and preferably include a code like A, M, P for reference later. In some experimental units (randomly allocated to different groups) treatments are simple, in others they are combinations, for example 4 or 40 mg ASA kg b.w. alone or in a combination with 5 mg Pycnogenol/kg b.w. give 4 combinations in total, if they are to be dosed twice a day, there are 8 different treatments. If all doses/combinations are administered at a morning, between 8.00 and 9.00, then the information about the time of dosing should be given in description of methods rather than here. Likewise, if the purpose of the experiment is to find the best time (morning or evening) to administer the drug at one selected dose, then the treatments are just times of administration and all details about the drug and its dose should be stated in "Methods". When we plan to use different drug doses we need to answer the question whether we want to compare the doses with each other or whether we want to compare both doses with the effect of doing nothing. In the latter case there is a third treatment, "do nothing", which is commonly called a *control*.

We should always decide whether "control" is really "a must", whether it is really needed. Or maybe "the necessary control" is rather our orthodox thinking of science? Be aware that in some experiments including the control group can be harmful; when it is already known that a given therapy is effective in ameliorating complications of a disease, then it would be unethical to run an experiment comparing a new tested therapy to a pure control (doing nothing). In such cases it would be much more reasonable to create a reference group (currently used therapy) instead of pure control. In medical practice, when experiments are performed on people, the statement "do nothing" should be replaced by a term "placebo", so that everyone involved thinks that something is being done (blinding).

- What are the methods to be used?  
There should be given technical details sufficient enough to enable other scientists to replicate our work. The description should include information on exactly how we apply the examined treatments to the chosen experimental units and what is to be done from then on until we collect all measurements.
- What are the experimental units?  
Experimental units are the elements of our examined population, which serve us to evaluate effects of the treatments. They need to be described in sufficient detail. For example, if there are 12 male Wistar rats sacrificed in the course of a 3-day experiment, then we have 12 experimental units. Otherwise, when we have 15 adult donors, and each of them is donating blood every day for 4 days, then we deal with 60 experimental units if we are able to change treatments every day. When choosing the experimental units we should keep in mind that it is not so important that they are as alike as possible, but it is extremely important that they are representative. The characteristics of the experimental units should be like a 'mirror image' of the characteristics of a general population the elements come from. We should realize that with experimental units unlike we risk that any within-group variation may conceal any differences between the treatments that we are trying to find, especially if our sample size is small. However, if they are not representative, we are not able to extrapolate any conclusions of our study.
- What are the observational units?  
The observational units are the objects on which we take measurements. In many situations they are the same as the experimental units. To specify what are the observational units we first need to declare which parameter(s) we are interested in measuring. If we have 15 streptozotocin-diabetic rats, given daily N-methylnicotinic acid in one dose for 8 weeks, and each animal is bleed once at the termination of an

experiment, then the observational units are the 15 rats. Otherwise, if each animal is bleed once a week, then the observational units are 120 rat-weeks. In our data sheet one row usually represents one observational unit.

- What measurements are to be recorded?

In our protocol it is necessary to write down everything to be recorded, for example:

- body mass in grams every 7 days, in the morning, after 12-hour fasting;
- blood glucose in mg/100 ml after 12-hour fasting;
- expression of platelet surface membrane P-selectin in samples of blood anticoagulated with 3.2% citrate from each of CAD patients from each group (each blood sample will be observational unit).

- What is the proposed design of our study?

The design of the study contains the description of experiment (e.g. 29 patients receive ASA, 34 patients receive Clopidogrel); in more complex designs we need to mention the use of blocks, Latin squares or Greek-Latin squares. Remember that in research we rely on inductive reasoning that moves from specific observations to broader generalizations and theories. The tool of induction is in fact “a trade mark” of science and research, and it may be so because it relies on numbers. The more replications our observations are based on, the stronger is our belief in a given generalization; we say that the lower is the risk that we may be wrong when stating a certain conclusion. We need, however, to justify the amount of replication undertaken in our study. If there is too much replication then we may waste time and money on our experiment. If our study involves laboratory animals that are to be sacrificed, it may be unethical to use too many. Otherwise, if there is too little replication, we risk that any real differences between treatments may be masked by the differences among the experimental units (within-group variability masks inter-group variability), so we may be unable to give any conclusions, and we are stretched to waste available resources (patients, animals; problem of ethics). To reasonably compromise we need to rely on one of the best research tools ever invented: the *a priori* estimation of a sample size.

- What is our plan of the study?

In a simpler approach, in an experiment with no blocks, we normally employ what is called a “complete randomization”. For example, to randomly select subjects/patients, whose names begin with letters B, C, D, F, J, K, P, R and S, to compare the treatments A and Z, we simply choose a random permutation (either by shuffling cards, tossing balls/envelopes, or from a computer) and apply it to the plan of our experiment. First we

systematically number our experimental units (patients)

1	2	3	4	5	6	7	8	9
B	C	D	F	J	K	P	R	S

and our suitable plan of action

1	2	3	4	5	6	7	8	9
A	A	A	A	Z	Z	Z	Z	Z

Then we overlay our (generated) random sequence of numbers 1 through 9, f.i. 1, 5, 6, 7, 9, 2, 8, 3, 4, to the systematic sequence of patients:

1	5	6	7	9	2	8	3	4
B	J	K	P	S	C	R	D	F

1	2	3	4	5	6	7	8	9
A	A	A	A	Z	Z	Z	Z	Z

- How do we propose to perform a statistical analysis? It is of crucial importance to consider the guidelines for the statistical analysis of data that we are supposed to collect. There are at least several rationales explaining why it is worthy to do it *a priori* (including those referring to the adjustment of experimental design according to how the collected data may be analyzed, e.g. ANOVA or regression models), before we collect the data. Simply, it is said: “if you cannot think how to analyze the data before their collection, do not waste your time and effort to collect it.” Of course, you may often end up accommodating or broadening your idea of statistical analysis, e.g. using *post hoc* multiple comparison tests as a consequence of the rejection of null hypothesis with ANOVA. In general, however, matching between what we planned and what we finally performed with respect to data analysis well testifies to how adequately we planned a whole experiment.

### Concluding remark – a single one

Randomization is more than a fad... and less than a revolution (in our thinking of experimentation).

### References

1. Armitage P, Berry G. Statistical Methods in Medical Research. 3<sup>rd</sup> edition. Blackwell, 1994.
2. Armitage P, Berry G, Matthews JN. Statistical Methods in Medical Research. 4<sup>th</sup> edition. Oxford: Blackwell Science, 2002.
3. Bland M. An Introduction to Medical Statistics. 3<sup>rd</sup> edition. Oxford Medical Publications, 2000.
4. Fog A. Pseudo-random number generators. 1999: <http://www.agner.org/random>.
5. Gentle JE. Random Number Generation and Monte Carlo Methods (Statistics and Computing). New York: Springer-Verlag, 2003.
6. Marsaglia G. DIEHARD: A battery of tests of randomness, 1997: <http://stat.fsu.edu/pub/diehard>.