



DOI: 10.2478/rmlm-2014-0046

Suggestions and recommendations on statistical analysis and research methodology

Septimiu Voidăzan¹, Cosmin Moldovan², Minodora Dobreanu³

*Editors, Revista Română de Medicină de Laborator
University of Medicine and Pharmacy, Tîrgu Mureş, Romania
1. Department of Epidemiology, 2. Department of Histology,
3. Department of Laboratory Medicine*

An inappropriate use of statistical methods may be associated with incorrect results, capable of leading to erroneous clinical practices, which can sometimes be a disadvantage for patients. False positive results will hinder the implementation of future studies that are used to restore the truth, ending up being difficult to be justified. On the other hand, false negative results may stimulate unnecessary research studies and a waste of resources. The problem is to determine what is good and what is bad, what is lofty and what is mediocre, what is probable and what is improbable. This is where statistics intervenes; if rigorously applied, it yields a precise value to the probability of occurrence of the studied effect.

The Romanian Review of Laboratory Medicine uses a working mechanism to monitor the statistical accuracy of scientific articles submitted to the editorial office [1].

Choosing the method of statistical analysis depends on the purpose and objectives of the study, the types of pursued variables and comparisons that are likely to be made. A well conducted statistical analysis should provide the opportunity to be firstly verified by reviewers, in order to decide whether the best statistical soft-

ware respectively the best statistical tests were decided upon so as not to mislead the reader.

There are situations identified in many papers evaluated from a statistical point of view, when the statistical software and the statistical tests applied in statistical processing are not mentioned, or there is no reference concerning the threshold of statistical significance compared to which the results of inferential statistics are interpreted. Usually, in statistics the statistical significance threshold of 0.05 is chosen, which means that a risk of 5/100 chances was decided in order to reject the null hypothesis, even if it is true.

The section from The Romanian Review of Laboratory Medicine that describes the policies of the journal and the instructions for authors contains aspects that must be observed strictly by the statistical analyses of data in studies [2-4].

For submitted papers to be published, it is compulsory to mention in the *Statistical analysis* section the statistical software and statistical tests which were used, respectively the established significance threshold. Concerning the statistical tests application method, differentiation between parametric and nonparametric tests

is required. The decision regarding the application of these tests concerns establishing, at the beginning of statistical processing, a descriptive statistical analysis in order to highlight a normal or Gaussian distribution, considering the indicators of central tendency (mean, median, mode) and those of dispersion (variance, standard deviation, coefficient of variation). The normal distribution is the domain where the values of a statistical variable are at a probability of 95% between the mean $\pm$ two standard deviations. In order to observe if a variable has a normal distribution, any statistical program can be used and the mean minus twice the standard deviation should not have negative values. Hence it is clear: parametric statistical tests can be applied to quantitative variables with normal distribution while non-parametric tests are to be applied to quantitative variables without normal distribution.

Descriptive statistics decides whether parametric or nonparametric tests are applied within the inferential or analytical statistics. Thus, the study must mention whether tests of normality have been applied at the beginning of the statistical processing and, implicitly, whether statistical processing was carried out or verified by a statistician. When there are doubts concerning statistical processing and suspicions about the correct application of the statistical tests and the correctness of the statistical significance, authors are requested to provide the database and possibly the statistical software they used so the correctness of the results can be verified.

Authors will also specify the modality of expressing data, such as: data were expressed as mean $\pm$ SD or mean $\pm$ SE, or median and interquartile range (IQR). Other expressions: *t* test as *t* (in italics), *F* test as *F* (in italics), chi square test as χ^2 (in Greek), related coefficient as *r* (in italics), degree of freedom as *v* (in Greek), sample number as *n* (in italics), probability as *P* (in italics), confidence interval *CI:95%*.

There are two major types of research themes which correspond to two broad study categories: descriptive studies, which describe the population distribution characteristics, and analytical studies that issue a judgment on a possible relationship between different factors that are studied.

In the *Material and method* section it is compulsory to mention the method of study and the type of study that has been performed: single-blind, double-blind, placebo, observational, experimental study, laboratory experiment, comparative experiments using drugs, clinical study, cohort, prospective, retrospective, case-control study. Study groups categorized based on inclusion and exclusion criteria have a statistically superior value if set prospectively.

The size of the study sample plays an important role in designing a study. If the number of subjects is insufficient, they were unnecessarily exposed to a state of discomfort, inconvenience or even risk, and an excessive number of subjects makes it possible for more people to be in the same potentially unpleasant position. In order to be considered representative, the sampling method which was used should be mentioned, as well as whether the qualitative (extracted randomly) or the quantitative notion (if a calculation method was used) was the one that was observed. The size of an appropriate sample according to the aim of the study should be calculated based on certain statistical methods which have to take into account several data assessments when an intermediary data analysis is required. It is not advisable to carry out intermediary statistical analyses during data collection, because if the results are statistically significant under the established working hypothesis, there is a tendency to end the study and prepare the paper for publication. If the results are "promising" but are not statistically significant, gathering additional data is compulsory. The intermediary statistical data analysis can be accompanied by

false positive results, and should therefore be avoided. The sample size and sampling methods defined while designing the study must be rigorously respected. This ensures that the results yielded by that representative sample are not produced by chance and can be extrapolated to the target population [5, 6].

If a questionnaire is used as a research tool, it should be mentioned whether it is validated and by what method. There are two essential qualities for a questionnaire: accuracy (ability to provide an accurate measurement of what should be measured), and reproducibility (the quality to provide a repeated identical measurement).

In descending order of the epidemiological studies, the most representative ones are the randomized clinical studies. From a methodological point of view they are considered the *gold standard* because they use randomization criteria in order to establish study groups or blinding. The methods of randomization and blinding should be well and thoroughly described by the authors in scientific papers. When referring to a randomized clinical study the authors must answer several questions: are the results of the study valid given that the allocation of subjects in the study was random?, was the randomization hidden?, were the groups similar at the beginning of the study?, was the research complete?, were the participants blind with respect to the intervention factor?, were the groups equally treated, except for the experimental intervention? [7].

Each epidemiological study is characterized by a design and must be strictly respected from the beginning.

All randomized clinical trials should include a flow diagram and authors should provide a completed randomized trial checklist (see CONSORT Flow Diagram and Checklist; and a trial protocol [8]. Studies of diagnostic accuracy must be reported according to STARD guidelines [9]; Observational studies (cohort, case-control, or cross-sectional designs) must be reported ac-

cording to the STROBE statement, and should be submitted with their protocols [10]; Genetic association studies must be reported according to STREGA guidelines [11]; Systematic reviews and meta-analyses must be reported according to PRISMA guidelines [12];

Generally studies are performed on humans, but there are situations when investigations on animals are carried out for experimental purposes. The most reliable studies on animals are those which use randomization to eliminate systematic differences between treatment groups and maintain comparability (assess the results in a blind manner) throughout the investigation, without knowing whether the animal has received the drug in question. Studies that do not report these measures are more likely to overestimate the effectiveness of interventions.

When conducting such experimental studies the working method, types of tests, types of reagents that were used, the equipment, the storage conditions for animals, food, investigations etc. have to be described.

In terms of perfecting an experimental study on animals, in order to analyze the data correctly and achieve the scientific objectives it is important, primarily for ethical but also for economic reasons, to use a minimum number of animals necessary for the study, but enough in order not to miss important biological effects and thus unnecessarily repeat the experiments. In the design phase of the study it is very important to ask the opinion of a statistician regarding the analysis of the working hypothesis developed by the researcher and the design of a proper experimental study methodology, so that all relevant information from the resulting data could be extracted. The statistician can provide advice on the study design, purpose, objectives, criteria for inclusion and exclusion, batch size estimation using power and sample size calculations.

The scientific paper should explain how the decision about the number of animals was tak-

en and provide details of any calculation used to decide sample size. Reporting the number of animals used in the study is essential, so that statistical and biological significance of the experimental results can be assessed and also necessary if the experimental methods must be repeated [13].

Using a small number of animals in the experimental studies is not only a requirement of the funding agencies around the world, but an ethical obligation as well. However, experiments performed on an insufficient number of animals are associated with reduced statistical power and little chance to detect significant differences between groups. In smaller studies, the positive predictive value is lower, the false-positive results are higher, which leads to unnecessary use of animals in further studies [14].

Additionally, at least in Romania experimental studies cannot be carried out with ease, because of the costs of reagents, investigations etc., which prevent the use of a large number of animals in studies.

Among the basic steps in preparing a study on animals, documents sent to the Ethics Commission, under international or institutional guidelines for the care and use of animals in research, have an important role.

The experimental and the control group should be carefully selected and all data should be fully recorded from the beginning of the study until the end (e.g. source of animals, species, sex, age, health, nutrition and growth conditions). Precise details of all performed procedures must be provided: drug doses, place and way of administration, type of anesthesia and analgesia, monitoring manner, surgical procedure, method of euthanasia, including details of any equipment used, suppliers, etc. The intervention factor (study drug, drug reference) must be randomly assigned to experimental groups, and the randomization procedure should be mentioned to avoid the effect of bias and implicitly false-pos-

itive results. Data collection must be performed identically using the same measuring instruments and calendar etc. When the blind method is used, the time should be noted: when, who, how. There are papers in the literature that provide information based on research carried out on animals. For example, the ARIVVE guides consist of a list of 20 items that briefly describe what all scientific publications that include animals in the research project should take into account, such as the number and the specific characteristics of the animals used, growth conditions, statistical and analytical methods, etc. [13, 15-17].

The statistician may prepare a basic database or provide advice in applying the most appropriate statistical tests and interpretations in order to extract as much information as possible in correlation with the aim and objectives set out in the study. The scientific paper contains details of the statistical methods used for each analysis. It is required to specify the unit of analysis for each set of data, such as reporting an entire group of animals or only anatomical elements, because there are often confusions regarding statistical processing. For example, a study performed on rabbits (80 eyes - 40 rabbits) in which the intra-ocular administration of a drug was monitored in the experimental group while the control group did not receive any medication. There are times when statistical analyses are performed on the number of animals or the number of eyes, using absolute frequency for each group included in the study instead of percentages (20/40, not 50%), primarily due to small samples. If an animal is not included in the statistical analysis, an explanation concerning the reason of exclusion must be provided. Direct analysis on animals, where loss of subjects or exclusion from the study may occur, may result in a more rapid decline of the sample, because an animal with an affected eye which does not comply with the study protocol will be removed from the scheme

even if the other eye is healthy. In contrast, direct analysis based on the number of eyes is not characterized by a drastic reduction in the sample because an affected eye which does not comply with the study protocol does not exclude the animal from the scheme, since the contralateral eye is still used. The results report for each analysis should include a measurement of precision (e.g. standard error or confidence interval: CI95%).

Statistical analysis uses the relationship between four statistical parameters which must be taken into consideration: sample size, type I and type II errors, the extent of the effect and the power of the study. The probability of committing a type I error is denoted by α and occurs when a difference between two study groups is stated, but in reality it does not exist. The probability of committing a type II error is denoted by β and occurs when a difference between the two study groups is not found, but in reality it exists. The power of the study is the probability of rejecting the null hypothesis when it is actually false, or in other words to highlight a difference when it actually exists. The power of the study is equal to $1 - \beta$ and depends on the sample size. It is widely accepted that the power of a study must be at least 80% to detect an acceptable level of the effect. If the power is weak, the researcher must decide whether or not he or she assumes the risk to conduct the study and the risk of having a negative result due to lack of power of the study [18, 19].

A statistically significant result does not provide information on the extent of the effect, therefore it does not necessarily mean that the effect is reliable. Also, it must be taken into account that statistic significance does not always translate into clinical significance.

The tendency to publish only positive results is another flaw in research conducted on animals. It sometimes provides a false impression of effectiveness, and at other times leads to an overstatement of efficacy of the performed experi-

ment. A study with negative results is less likely to be published than one with positive results.

For the purpose of any study, whether carried out on humans or animals, the limits of the study should be mentioned and discussed, with reference to the sources of bias or the inaccuracies that may be associated with the results.

In conclusion, the dissemination of research results is the engine that enhances scientific progress. Concerning result evaluation, it is important for scientific communities that studies conducted both on humans and animals transparently provide sufficient detail on the design, conduct and analysis of experiments. Awareness of the importance of a rigorous study design will increase the quality of studies and thus accelerate scientific progress.

References

1. Adrian Man, Cosmin Moldovan, Minodora Dobrea-nu, Major editorial changes of Romanian Review of Laboratory Medicine. *Rev Romana Med Lab.* 2013;21(4):377–80. DOI: 10.2478/rrlm-2013-0047
2. Romanian Review of Laboratory Medicine/Versita [Internet]. [cited 2013 Nov 10]. Available from: <http://versita.com/serial/romanian-review-of-laboratory-medicine/>
3. Revista română de medicină de laborator [Internet]. [cited 2013 Dec 1]. Available from: http://rrml.ro/instructiuni_autori.php
4. license_to_publish.pdf [Internet]. [cited 2013 Nov 23]. Available from: http://rrml.ro/instr_autori/license_to_publish.pdf
5. Dienes Z. Bayesian versus orthodox statistics: which side are you on? *Perspect Psychol Sci.* 2011;6:274–290. DOI: 10.1177/1745691611406920
6. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci.* 2011 Nov;22(11):1359–66. DOI: 10.1177/0956797611417632
7. Straus S, Richardson S, Glasziou P, Haynes B. Evidence based medicine. How to practice and teach EBM. 3rd Edition, Elsevier 2005, p.115.
8. <http://www.consort-statement.org>
9. <http://www.stard-statement.org>
10. <http://www.strobe-statement.org>
11. <http://www.medicine.uottawa.ca>

12. <http://www.prisma-statement.org>
13. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLoS Biol. 2010;8(6):e1000412. DOI: 10.1371/journal.pbio.1000412
14. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2:e124. DOI: 10.1371/journal.pmed.0020124
15. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW et al. A call for transparent reporting to optimize the predictive value of preclinical research. Nature. 2012 Oct 11;490(7419):187-91. DOI: 10.1038/nature11556
16. Sena E, van der Worp HB, Howells D, Macleod M. How can we improve the pre-clinical development of drugs for stroke? Trends Neurosci. 2007;30:433-9. DOI: 10.1016/j.tins.2007.06.009
17. Festing MF, Altman DG. Guidelines for the design and statistical analysis of experiments using laboratory animals. ILAR J. 2002;43(4):244-58. DOI: 10.1093/ilar.43.4.244
18. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition. Erlbaum, Hillsdale, NJ. 1988.
19. Nakagawa S, Foster TM. The case against retrospective statistical power analyses with an introduction to power analysis. Acta Ethologica. 2004;7(2):103-8. DOI: 10.1007/s10211-004-0095-z