

EDITORIAL

doi: 10.5455/medarh.2019.73.298-302

MED ARCH. 2019 OCT; 73(5): 298-302

RECEIVED: SEP 22, 2019 | ACCEPTED: OCT 15, 2019

¹Faculty of Medical Sciences, University of Kragujevac, Kragujevac, Serbia

²Bosnalijek, Sarajevo, Sarajevo, Bosnia and Herzegovina

³Academy of medical sciences of Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina

Corresponding author: Professor Slobodan M. Jankovic, MD, MSc, DSc, Prim. University of Kragujevac, Faculty of Medical Sciences, Svetozara Markovića Street, 69, Kragujevac, 34000, Serbia. E-mail: slobnera@gmail.com. ORCID ID: <https://www.orcid.org/0000-0002-1519-8828>.

Evaluation of Published Preclinical Experimental Studies in Medicine: Methodology Issues

Slobodan M. Jankovic¹, Belma Kapo², Aziz Sukalo², Izet Masic³

ABSTRACT

Introduction: Inappropriate design of experimental studies in medicine inevitably leads to inaccurate or false results, which serve as basis for erroneous and biased conclusions. **Aim.** The aim of our study was to investigate prevalence of implementing basic principles of experimental design (local control, replication and randomization) in preclinical experimental studies, performed either on animals in vivo, or animal/human material in vitro. **Material and Methods.** Preclinical experimental studies were retrieved from the PubMed database, and the sample for analysis was randomly chosen from the retrieved publications. Implementation rate of basic experimental research principles (local control, randomization and replication) was established by careful reading of the sampled publications and their checking against predefined criteria. **Results.** Our study showed that only a minority of experimental preclinical studies had basic principles of design completely implemented (7%), while implementation rate of single aspects of appropriate experimental design varied from as low as 9% to maximum 86%. Average impact factor of the surveyed studies was high, and publication date relatively recent, suggesting generalizability of our results to highly ranked contemporary journals. **Conclusion.** Prevalence of experimental preclinical studies that did not implement completely basic principles of research design is high, raising suspicion to validity of their results. If incorrect and biased, results of published studies may mislead authors of future studies and cause conduction of fruitless research that will waste precious resources.

Keywords: randomization, control experiments, replication, internal validity.

1. INTRODUCTION

Inappropriate design of experimental studies in medicine inevitably leads to inaccurate or false results, which serve as basis for erroneous and biased conclusions (1). Although numerous attempts were made in the past to prevent errors in research design, like establishing guidelines for experimental studies (2) or teaching experimental design at postgraduate studies (3), evidence shows that some of the basic principles of experimental research design are still not implemented in more than half of the studies published in medical journals (4). There are three basic principles of experimental design that guarantee reliability of the results: having appropriate negative and positive controls for treatment or a factor that is tested, replicating experiments on independent experimental units sufficient number of times and randomly assigning a treatment (or factor) that is tested and control treatment (or factor) to experimental units (5). Failure to acknowledge and implement these principles when planning

a study usually causes production of false positive experimental results, which are rather consequence of uncontrolled factors, like concomitant conditions or maturation of experimental units, than of the treatment (or a factor) that is actually tested (6).

2. AIM

The aim of our study was to investigate prevalence of implementing basic principles of experimental design (local control, replication and randomization) in preclinical experimental studies, performed either on animals in vivo, or animal/human material in vitro.

3. MATERIAL AND METHODS

The studies were retrieved for analysis from the PubMed database. The following inclusion criteria defined the pool of the studies from which the study sample was extracted: journal article, original experimental study, animal study, in vitro study and full text availability. The exclusion criteria were: review articles, clinical trials of phase I-IV, cohort

© 2019 Slobodan M. Jankovic, Belma Kapo, Aziz Sukalo, Izet Masic

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Requirement	Satisfied n (%)	Not satisfied n (%)	Unclear n (%)	Not applicable n (%)
Sample size reported for the experiment?	26 (60%)	17 (40%)	-	-
Number of observations reported for the experiment	33 (77%)	10 (23%)	-	-
Value of test statistics, exact p value and degrees of freedom reported	6 (14%)	37 (86%)	-	-
Error bars correspond to the analysis (i.e. standard error is based on number of independent observations)	19 (44%)	6 (14%)	12 (28%)	6 (14%)
Only independent observations were taken into account for statistical tests	19 (44%)	4 (9%)	20 (47%)	-
Is there negative control?	32 (74%)	11 (26%)	-	-
Was positive control necessary, and if so, was it used?	20 (47%)	19 (44%)	-	4 (9%)
Were treatments randomly allocated to experimental units?	4 (9%)	28 (65%)	2 (5%)	9 (21%)
Number of citations: mean, standard deviation, median, interquartile range	28.6 ± 37.2; 12.0; 29.0			
Time passed from the publication (years)	12.4 ± 10.9; 9.0; 13			

Table 1. Results of the survey of the experimental studies (n = 43)

studies, case control studies and cross-sectional studies. The following search strategy was used to implement inclusion and exclusion criteria and select the pool of the studies for further analysis: ("animals"[MeSH Terms:no-exp] OR animal[All Fields]) AND study[All Fields] OR ("in vitro techniques"[MeSH Terms] OR ("vitro"[All Fields] AND "techniques"[All Fields]) OR "in vitro techniques"[All Fields] OR "vitro"[All Fields] OR "in vitro"[All Fields]) AND study[All Fields] NOT ("review"[Publication Type] OR "review literature as topic"[MeSH Terms] OR "review"[All Fields]) AND (Journal Article[ptyp] AND "loattrfree full text"[sb]).

Size of the study sample (n=43) was calculated on the basis of the following assumptions: rate of inappropriate research design 0.5 (4) and width of the 95% confidence interval ± 0.15 . The formula $n = (1.96)^2 \times 4 \times p \times (1-p) / d^2$ was used for the calculation, where „n“ is the sample size, „p“ probability of inappropriate research design and „d“ width of the confidence interval (7). Since the studies retrieved by the abovementioned search strategy were numbered orderly in the PubMed database, the study sample of 43 studies was extracted by simple randomization technique, activating for 43 times random number generator in Excel, using formula RANDBETWEEN(1;666;342).

The extracted studies were analyzed for internal methodological validity, checking whether basic principles of correct experimental design (replication, control and randomization) were implemented. For the purpose of this analysis, the checklist with 8 questions was prepared, as shown in the Table 1. The results of the analysis were tabulated and described by rates and percentages when categorical, and by means, standard deviations, medians and interquartile ranges, if continuous.

4. RESULTS

In total 43 journal articles were retrieved randomly from pool of 666,342 articles in the PubMed database defined by the inclusion and exclusion criteria, and then analyzed according to predefined criteria of research design quality. Average impact factor of the journals (for the years when the articles were published) was 3.9739 ± 1.9125 , median impact factor was 3.7490, and interquartile range 2.240. Compliance of the articles with the

criteria, average number of citations per article and average time elapsed from the publication of the articles are shown in the Table 1. Only three of the analyzed studies (7.0%) had all basic principles of experimental design completely implemented.

Number of satisfied criteria per study was not correlated either with journal impact factor (Spearman's rho = 0.058, $p = 0.710$) or with number of citations (Spearman's rho = -0.254, $p = 0.100$). The time elapsed from the publication also was not correlated with the number of satisfied criteria per study (Spearman's rho = -0.227, $p = 0.144$).

5. DISCUSSION

Our study showed that only a minority of experimental preclinical studies had basic principles of design completely implemented (7%), while implementation rate of single aspects of appropriate experimental design varied from as low as 9% to maximum 86%. Average impact factor of the surveyed studies was high, and publication date relatively recent, suggesting generalizability of our results to highly ranked contemporary journals. Prevalence of certain aspects of inappropriate design in our study was similar to values reported by other studies, especially in regard to lack of randomization, which was observed in 70% of studies from our sample and in 87% of studies surveyed by Kilkenney et al (8). A number of the authors of experimental studies on animals, human cells, or tissues are misled by superficial similarity between the experimental units, derived from their common origin (the same cell line, the same clone of animals, the same species from the same breeding line, etc.), and may wrongly assume that they are completely the same. However, even identical twins are not completely identical, as many external factors with shape them differently, so randomization is always necessary in experimental studies, regardless of the similarities between the experimental units (9).

While necessity of having local control in their experiments was understood by authors of majority of analyzed studies, the replication issue remained obtunded, and difference between true replication (repeating experiments on independent experimental units) and pseudo replication (repeating experiments on the same

Evaluation of Published Preclinical Experimental Studies in Medicine

Study	Sample size reported for the experiment?	Number of observations reported for the experiment	Value of test statistics, exact p value and degrees of freedom reported	Error bars correspond to the analysis (i.e. standard error is based on number of independent observations)	Only independent observations were taken into account for statistical tests	Is there negative control?	Was positive control necessary, and if so, was it used?	Were treatments randomly allocated to experimental units?	Number of citations
Study 1	no	no	no	no	no	no	no	no	71
Study 2	yes	no	no	no	It is not clear	yes	no	no	25
Study 3	yes	no	no	no	unclear	no	no	no	11
Study 4	yes	yes	yes	Not shown	yes	yes	yes	no	64
Study 5	yes	yes	no	unclear	unclear	yes	no	no	7
Study 6	no	yes	no	yes	unclear	yes	no	no	23
Study 7	no	yes	no	no	unclear	no	yes	no	1
Study 8	yes	yes	no	yes	yes	yes	no	Yes, but not explained how	5
Study 9	no	no	yes	Not clear	Not clear	yes	no	no	36
Study 10	no	yes	no	no	yes	yes	yes	Mentioned, but not explained	8
Study 11	yes	yes	no	yes	Not clear	yes	no	no	3
Study 12	yes	yes	no	yes	yes	no	yes	yes	4
Study 13	no	yes	no	yes	yes	no	no	Not applicable	33
Study 14	no	yes	no	Not shown	no	no	no	Not applicable	122
Study 15	yes	yes	no	yes	yes	yes	yes	yes	60
Study 16	yes	yes	no	Not applicable	Not clear	no	yes	Not applicable	4
Study 17	yes	yes	no	yes	Not clear	yes	no	Not applicable	12
Study 18	yes	yes	yes	yes	yes	yes	yes	no	153
Study 19	No	yes	No	Not clear	No	Yes	Yes	no	8
Study 20	yes	yes	yes	yes	yes	yes	yes	Not applicable	12
Study 21	yes	yes	no	Not applicable	yes	no	yes	no	5
Study 22	yes	yes	no	Not applicable	yes	yes	no	Not applicable	9
Study 23	yes	yes	no	yes	yes	no	yes	yes	2
Study 24	yes	yes	no	Not clear	yes	yes	no	no	6
Study 25	no	no	no	Not applicable	Not clear	no	no	no	19
Study 26	No	yes	no	yes	yes	yes	yes	no	11
Study 27	yes	yes	no	yes	yes	yes	no	Not applicable	9
Study 28	no	no	no	Not applicable	Not clear	no	no	Not applicable	159
Study 29	yes	no	no	yes	Not clear	yes	yes	no	20
Study 30	yes	no	no	yes	yes	yes	no	no	27
Study 31	yes	No	no	Not clear	Not clear	yes	no	no	8
Study 32	yes	yes	no	yes	yes	yes	yes	no	3
Study 33	yes	yes	yes	yes	yes	yes	Not applicable	yes	7
Study 34	No	Yes	no	yes	Not clear	yes	yes	no	1
Study 35	yes	yes	no	yes	Not clear	yes	yes	no	30
Study 36	no	yes	no	no	no	yes	yes	no	50
Study 37	no	yes	no	yes	yes	yes	yes	no	13
Study 38	yes	yes	yes	Not applicable	yes	yes	Not applicable	Not applicable	31

Study	Sample size reported for the experiment?	Number of observations reported for the experiment	Value of test statistics, exact p value and degrees of freedom reported	Error bars correspond to the analysis (i.e. standard error is based on number of independent observations)	Only independent observations were taken into account for statistical tests	Is there negative control?	Was positive control necessary, and if so, was it used?	Were treatments randomly allocated to experimental units?	Number of citations
Study 39	no	yes	no	Not clear	Nor clear	yes	No	no	41
Study 40	yes	yes	no	Not clear	Not clear	yes	yes	no	7
Study 41	yes	yes	no	Not clear	Not clear	yes	Not applicable	no	49
Study 42	no	yes	no	Not clear	Not clear	yes	Not applicable	no	22
Study 43	No	No	no	Not clear	Not clear	yes	yes	no	40

Table 2. Check list of 43 published papers used for assessment of the frequency of un-adequated study design

experimental unit) was not appreciated by majority. Pseudo replication leads to inappropriate testing of hypothesis (because statistical tests for testing difference between groups assume independence of experimental units) and to false precision, as improved estimate after repeating measurements on the same experimental unit just gives more precise results for that experimental unit, and not for the population that is investigated (10-13). Pseudo replication can also undermine the conclusions of a statistical analysis, and it would be easier to detect if the sample size, degrees of freedom, the test statistic, and precise p-values are reported. This information should be a requirement for all publications.

The articles we analyzed in this study were highly cited regardless of their methodological shortcomings and possibly wrong conclusions, that may lead to erroneous assumptions when designing future studies and unnecessary wasting of research resources (14). Seven threats to the internal validity of experiments were discussed by Donald T. Campbell in his classic 1957 article: history, maturation, testing, instrument decay, statistical regression, selection, and mortality. These concepts are said to be threats to the internal validity of experiments because they pose alternate explanations for the apparent causal relationship between the independent variable and dependent variable of an experiment if they are not adequately controlled. Unlike with observational studies, experimental design is based on elimination of confusing variables with inclusion/exclusion criteria and on control of extraneous factors by setup of the experiment which should include randomization and local control. It is critical that experimental setup excludes extraneous influences, because they are not taken into account during statistical analysis, and may bias the results. If basic principles of experimental design are not implemented, extraneous variables will not be controlled properly, and the observed effects on experimental model may not be consequence of the tested treatment or factor, but of the extraneous variables themselves (15).

Widespread failure to comply with basic rules of experimental design also led to crisis in reproducibility of experimental results. Many so-called breakthroughs in experimental science turned out to be spurious and false

when independent study groups tried to repeat experiments described in published papers.

Some authors believe that majority of published experimental results will not stand the test of time (16), because numerous authors all over the world do not adhere to good experimental practice being under pressure to “publish or perish.” Some of measures that could improve the situation are: insisting on standards of data presentation, publication of negative results in scientific journals, and changes in principles of funding research that would prevent making profit on just having publications without any real impact on science and healthcare (17, 18).

Limitations of the study

The results of our study are limited to only one database (PubMed) having journals that are on average ranked highly than journals in some other databases with less strict inclusion criteria. Therefore, our results could underestimate the problem of inadequate experimental design, and should be interpreted with caution. Besides, not all published papers had enough data presented to allow for complete estimate of methodological issues.

6. CONCLUSION

Prevalence of experimental preclinical studies that did not implement completely basic principles of research design is high, raising suspicion to validity of their results. If incorrect and biased, results of published studies may mislead authors of future studies and cause conduction of fruitless research that will waste precious resources.

- **Author's contribution:** S.M.J, B.K, A.S, I.M. gave a substantial contribution to the conception and design of the work. S.M.J, and I.M. gave a substantial contribution of data. S.M.J, and B.K, gave a substantial contribution to the acquisition, analysis, or interpretation of data for the work. S.M.J. and I.M. had a part in article preparing for drafting or revising it critically for important intellectual content. All authors gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
- **Conflict of interest:** not declared.
- **Financial support:** nil.

REFERENCES

1. Auger JP, Chuzeville S, Roy D, Mathieu-Denoncourt A, Xu J, Grenier D, et al. The bias of experimental design, including strain background, in the determination of critical *Streptococcus suis* serotype 2 virulence factors. *PloS One*. 2017; 12(7): e0181920.
2. Moorhead JE, Rao PV, Anusavice KJ. Guidelines for experimental studies. *Dent Mater Off Publ Acad Dent Mater*. 1994 Jan; 10(1): 45-51.
3. Morsink MC, Dukers DF. Teaching neurophysiology, neuropharmacology, and experimental design using animal models of psychiatric and neurological disorders. *Adv Physiol Educ*. 2009 Mar; 33(1): 46-52.
4. Vesterinen HM, Vesterinen HV, Egan K, Deister A, Schlattmann P, Macleod MR, et al. Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the *Journal of Cerebral Blood Flow and Metabolism*. *J Int Soc Cereb Blood Flow Metab*. 2011 Apr; 31(4): 1064-1072.
5. Fry DJ. Teaching experimental design. *ILAR J*. 2014; 55(3): 457-471.
6. Knapp TR. Why Is the One-Group Pretest-Posttest Design Still Used? *Clin Nurs Res*. 2016 Oct; 25(5): 467-472.
7. Janković S. *Dizajn istraživanja*. 1st ed. Medrat, Kragujevac, Serbia, 2016.
8. Kilkenny C, Parsons N, Kadyszewski E, Festing MFW, Cuthill IC, Fry D, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PloS One*. 2009 Nov 30; 4(11): e7824.
9. Hoerauf JM, Moss AF, Fernandez-Bustamante A, Bartels K. Study Design Rigor in Animal-Experimental Research Published in *Anesthesia Journals*. *Anesth Analg*. 2018; 126(1): 217-222.
10. Lazic SE. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? *BMC Neurosci*. 2010 Jan 14; 11: 5.
11. Flannelly KJ, Flannelly LT, Jankowski KRB. Threats to the Internal Validity of Experimental and Quasi-Experimental Research in Healthcare. *J Health Care Chaplain*. 2018 Sep; 24(3): 107-130.
12. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull*. 1957 Jul; 54(4): 297-312.
13. Begley CG, Ioannidis JPA. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res*. 2015 Jan 2; 116(1): 116-126.
14. Pusztai L, Hatzis C, Andre F. Reproducibility of research and preclinical validation: problems and solutions. *Nat Rev Clin Oncol*. 2013; 10(12): 720-724.
15. Masic I, Kujundzic E. *Science Editing in Biomedicine and Humanities*. Avicena. Sarajevo, 2013.
16. Masic I, Jakovljevic M, Sinanovic O, Gajovic S, Jankovic MS, et al. The Second Mediterranean Seminar on Science Writing, Editing and Publishing (SWEP 2018), Sarajevo, December 8th, 2018. *Acta Inform Med*. 2018 Dec; 26(4): 284-296. doi: 10.5455/aim.2018.26.284-296.
17. Masic I. Medical Publication and Scientometrics. *Journal of Research in Medical Sciences*. 2013 Jun; 18(6): 624-630.
18. Masic I, Jankovic MS, Begic E. PhD Students and the Most Frequent Mistakes During Data Interpretation by Statistical Analysis Softwares. *Proceedings of ICITHM 2019*. Athens, Greece, 5-7 July 2019.