

Biostatistics in allergy – recommendations for authors

Allergy is the official journal of the European Academy of Allergy and Clinical Immunology (EAACI), as well as the number one in the medical field of allergology: 2021 Journal Citation Reports (Clarivate Analytics): 1/28 (Allergy).¹ One of the factors that has contributed to the prestige of Allergy is the high quality of the published research results. This article has been written with the aim of further raising the standard of the journal *Allergy*, namely by providing authors with recommendations to be implemented before they decide to submit an article. They were taken into account on the basis of numerous statistical reviews, which made it possible to identify the most frequent errors made by the authors in the processing of the obtained research results. Efforts should be made to improve the quality of statistical analysis worldwide, because at the moment it leaves much to be desired.^{2,3}

The first recommendation is that authors shall include in the cover letter information on the co-authors responsible for performing the statistical analysis. If an expert biostatistician had performed the analysis, this fact would have been noted. And when the journal editors have doubts about the quality of the research results analyzed, a statistical review is necessary.

The second recommendation has to do with the detailed description of the statistical tests used by the authors. A common error in submitted/accepted manuscripts, is a too general description of the statistical tests used. An example is a sentence written in the style of: "Statistical analysis used the *t*-student test, Mann-Whitney *U*-test and two-factor analysis of variance. This is often the first signal that may indicate uncertainty about the quality of the research results presented later in the publication. The purpose of the statistical tests used should be presented in a more precise way. If the authors mention a series of different tests, it seems necessary to specify in detail for what specific purpose they were used. A correctly worded sentence might read as follows: "To test for statistically significant differences between the allergic rhinitis group and the control group, the Student's *t*-test for independent samples or its non-parametric equivalent, the Mann-Whitney *U*-test, was used. The non-parametric equivalent was specific to the eosinophilia level type variable due to significant deviations from the normal distribution, as supported by the Shapiro-Wilk test result. The use of the Mann Whitney *U* test was also related to the unequal nature of the groups of people being compared. Another rather common example concerns the failure to describe further how to analyze a statistically significant interaction of variables in a multivariate analysis of variance. When choosing a particular post-hoc

test, conducting a simple regression analysis, authors very often do not take into account in related articles such assumptions as, for example, homogeneity of variance, equality of groups compared, correlation between predictors included in the regression analysis, homoscedasticity, and influence of outliers on the obtained results. One common example encountered in the course of conducting reviews is the presentation of, for example, the results of a correlation analysis, and the graphs provided show outliers which have a certain impact on the result obtained. To summarize this type of recommendation, submitted articles should include a detailed description of the purpose as well as the assumptions explored specific to the statistical tests used. This would significantly increase the value of manuscripts published in Allergy, as well as reduce the problem of reviewers' frequent doubts about the quality of the statistical analysis performed.

The next recommendation contains one, but very important, tip to increase the quality of published research results. Unfortunately for today, this is an under-emphasized aspect, namely the size of the effect. The *p*-value alone is definitely not enough. Effect size is a quantitative measure of the strength of a phenomenon calculated from data. The most commonly calculated measures of effect size/strengths include Cohen's *d*, Hedges' *g*, eta-square, *F*i coefficient, Cramer's *F*i, and Glass' *r*_g rank binary correlation coefficient. The effect size allows an assessment of the real, practical significance of the result, while the *p*-value allows statistical significance of the result. It also allows the results of multiple studies to be compared in meta-analyses. It is recommended that authors read articles in which the authors justify in detail why this *p*-value is not sufficient: "Using Effect Size-or Why the *p* Value Is Not Enough".^{4,5} Below are worthwhile quotes for the authors from experts pointing out the importance of effect size:

Statistical significance is the least interesting thing about the results. You should describe the results in terms of measures of magnitude—not just, does a treatment affect people, but how much does it affect them

(Gene V. Glass)⁶

The primary product of a research inquiry is one or more measures of effect size, not *p* values

(Jacob Cohen)⁷

TABLE 1 Key statistical recommendations recommended to authors planning to submit an article to Allergy

Statistical analysis	Recommendation
Persons responsible for statistical analysis	Including in the cover letter and manuscript information on the co-authors responsible for carrying out the statistical analysis (e.g., biostatisticians)
Statistical tests used	Including in the article a detailed description of the application of each statistical test—not a total in the style of, for example, the Mann-Whitney <i>U</i> test was used to see if there were differences between two independent groups of people.
Assumptions of statistical tests	Describing the assumptions of the statistical tests used and the associated results (normality of distribution, homogeneity of variance, sphericity of variance, effect of outliers/extreme cases, for example, on the result of correlation analysis, rationale for choosing a particular post-hoc test, correlations of predictors when conducting regression analysis, homoskedasticity).
Statistical test results	Recording the results of statistical tests according to scientific standards, for example, Mann-Whitney <i>U</i> test: $U = 20$; $p = 0.03$.
Effect size	Where there are statistically significant differences/dependencies, examining the effect size as measured by the coefficients appropriate to the individual statistical tests (Cohen's <i>d</i> , Hedges' <i>g</i> , eta-square, <i>Fi</i> coefficient, Cramer's <i>Fi</i> , and Glass' <i>rg</i> rank binary correlation coefficient). Objective—to learn about assessing the real, practical meaning of the outcome.
Descriptive statistics	The inclusion of descriptive statistics appropriate for the various statistical tests in the statistical analysis performed. Example suggestion: when using the Mann-Whitney <i>U</i> -test, do not just focus on the mean and standard deviation, but also include the median.
Clarity of the description of the statistical tests used.	If advanced statistical tests are used, describing in the article in simple language that is clear to the reader what the statistical tests used are about—not writing 2 pages on the subject in the article, but so that every reader understands what is involved.
Results in the abstract	Including in the abstract the most important results obtained—not writing in general style, for example, giving only a single indication of the existence of differences between two groups.
Open-access data since	Including publicly available data when sending the manuscript in order to achieve reproducibility of the results obtained. This would allow an interested party to download the raw results to check them without a request specifically created for this purpose.
Other aspects	Description and significance of missing data, clear indication under tables/charts of the statistical tests used, full information on the statistical package used, consideration of factors that may affect the results, such as gender, age, and comorbidities.

One definition of effect size also reads as follows: *Effect size is a quantitative representation of the magnitude of some phenomenon needed to address the issue being addressed.*⁸ The effect size allows the results of different procedures, or experiments with different group sizes, to be compared. It makes it possible to determine the sample size, as well as to examine the strength of the relationship under study, not just whether it is statistically significant or not. The authors focusing solely on the statistical significance of results may report results that are not relevant in practice. For this reason, one of the important in recommendation statistics, is the effect size, through which more practical information can be extracted.

The next recommendation is for the authors to describe the more advanced statistical tests that were used. In the absence of an explanation of what the tests used are, it may be difficult for the reader/reviewer to understand the research results analyzed. Including an additional few sentences in the text to explain to the reader the meaning of the advanced statistical test used, increases the clarity of the published articles. There are also situations when the description is too long regarding the tests used, while the article deals with a completely different topic. For this reason, this recommendation indicates that it should be a simple, reliable message, so that, the person reading the article can understand it properly.

Other key statistical recommendations that authors should keep in mind when submitting articles to Allergy include using descriptive statistics appropriate to the statistical tests, explaining any missing data that may appear, clearly labeling in the results the statistical tests used, and including information about the statistical package used. Unfortunately, it often happens that authors mention several tests they used, while in the presented results, it is not clear where and which test allowed to verify the presented results. Other errors include failure to calculate descriptive statistics appropriate for non-parametric equivalents of statistical tests. An example of this is to calculate only the arithmetic mean using the Mann-Whitney *U* test. Another not infrequent error that appears in the statistical analysis performed by the authors is the failure to take into account factors that may affect the results obtained, such as gender and comorbidities. For this reason, before authors decide to submit an article to Allergy, they should analyze at the outset whether such factors exist. As an example, one could mention the prevalence of asthma, where factors such as, for example, gender, age, and not just the entire patient group should be included in the analysis. What is noteworthy, and worth bearing in mind if you are thinking of increasing the reliability of the results of your research, is to include in the article a record of the results of statistical tests according to scientific

standards. Such an example is the result of the Mann–Whitney U test: $U = 23$; $p = 0.02$.

A final recommendation is to include the most important results in the abstract of the submitted manuscript. The authors often do not include in the abstract the most relevant results they obtained from their statistical analysis, but only sentences along the lines of: “There is an association between the level of eosinophilia and the severity of the disease entity studied” Specifying the key findings in the abstract is an essential next step to enable adequate conclusions to be drawn. The advantage of this type of recommendation is a more accurate depiction of the innovation associated with the published article, and thus affects the prestige of the journal.

To conclude this article, the recommended recommendations are intended to improve future submissions by authors to *Allergy*. Table 1 lists the most important aspects that authors should consider when submitting a manuscript to *Allergy*. The implementation of these recommendations aims to reduce the doubts of the editorial board members about the quality of the analysis, to increase the transparency of the published research results and thus to maintain the high level of the journal *Allergy*.

FUNDING INFORMATION

None.

CONFLICT OF INTEREST

The author declares no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Michał Ordak 

Department of Pharmacodynamics, Centre for Preclinical,
Research and Technology (CePT), Medical University of Warsaw,
Warsaw, Poland

Correspondence

Michał Ordak, Department of Pharmacodynamics, Centre
for Preclinical Research and Technology (CePT), Medical
University of Warsaw, 1B Banacha Street, Warsaw 02-097,
Poland.

Email: mordak@wum.edu.pl

ORCID

Michał Ordak  <https://orcid.org/0000-0001-5652-4397>

REFERENCES

1. <https://jcr.clarivate.com/jcr/browse-journals>
2. Ordak M. COVID-19 research: quality of biostatistics. *Arch Med Sci*. 2021;18(1):257-259.
3. Diong J, Butler AA, Gandeia SC, Héroux ME. Poor statistical reporting, inadequate data presentation and spin persist despite editorial advice. *PLoS One*. 2018;13(8):e0202121.
4. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. *J Grad Med Educ*. 2012;4(3):279-282.
5. Solla F, Tran A, Bertoncelli D, Musoff C, Bertoncelli CM. Why a p -value is not enough. *Clin Spine Surg*. 2018;31(9):385-388.
6. Kline RB. *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*. American Psychological Association; 2004:95.
7. Cohen J. Things I have learned (so far). *Am Psychol*. 1990;45:1304-1312.
8. Kelley K, Preacher KJ. On effect size. *Psychol Methods*. 2012;17(2):137-152.