

---

*Statistical Accuracy in Rheumatology Research*

**Ilke Coskun Benlidayi**

**Mediterr J Rheumatol 2019;30(4):207-15**



E-ISSN: 2529-198X



## Statistical Accuracy in Rheumatology Research

Ilke Coskun Benlidayi 

Cukurova University Faculty of Medicine, Department of Physical Medicine and Rehabilitation, Adana, Turkey

### ABSTRACT

Science and related research activities are always subject to errors. Statistical accuracy is necessary in order to overcome erroneous researching and reporting practices. The present article aimed to review the current literature on statistical errors in clinical medicine articles, and to provide rheumatologists with basic recommendations regarding the use of common statistical methods in research articles. With this purpose, PubMed/MEDLINE and Web of Science databases were searched by using relevant keywords. Data so far indicate that statistical errors are common in published articles from several disciplines of medicine. Statistics is the key element of any research activity, thus, implementing statistics at each step (hypothesis development, study design, sampling/data collection, data analysis, presentation) of every research is mandatory. In this regard, awareness of common statistical errors, basic knowledge on statistical methodology and consulting an expert in biostatistics from the beginning of the research process would be of value for rheumatologists.

*Mediterr J Rheumatol* 2019;30(4):207-15

<https://doi.org/10.31138/mjr.30.4.207>

Article Submitted: 15 Jul 2019; Revised Form: 11 Nov 2019; Article Accepted: 14 Nov 2019

**Keywords:** data accuracy, data analysis, data collection, data interpretation, rheumatology, medicine, statistics

### INTRODUCTION

Rheumatology is an ever-evolving area of medicine bearing a high research potential. Researching is necessary to build up evidence regarding the uncertain or unsolved topics in rheumatology. Publishing the results of these research activities is mandatory for delivering scientific knowledge among other researchers, health-care professionals, and patients. However, the accuracy, reliability and overall quality of publications are of great importance. Statistical accuracy is the key element for a reliable and quality research article. Therefore, it is of par-

amount importance to ascertain statistical appropriateness from the very beginning of a study until the end of the whole process. Science has always been subject to error.<sup>1</sup> Publishing erroneous results derived from inadequately managed research projects may mislead future researchers as well as physicians who intend to use these results in daily clinical practice. In this regard, mentorship is of importance for guiding research naive trainees to manage research projects. A mentor is expected to assist a mentee on how to go about starting a research project, using search engines, discussing the whole process with a biostatistics expert and taking the project to completion. On the other hand, awareness about common errors in scientific research is also essential. Over the past years, authors have focused on the importance of statistical correctness in medical research articles from different disciplines.<sup>2-14</sup> However, data on

#### Corresponding Author:

Ilke Coskun Benlidayi, MD  
Associate Professor  
Department of Physical Medicine and  
Rehabilitation  
Faculty of Medicine  
Cukurova University, Adana  
Turkey  
Tel.: + 90 538 545 39 37  
Fax: + 90 322 338 64 29  
E-mail: icbenlidayi@hotmail.com

statistical errors and/or accuracy in rheumatology articles are scarce.<sup>12,15</sup> Given the lack in the literature, the present article aimed to provide recommendations for rheumatologists to ascertain statistical appropriateness in research papers by highlighting common statistical errors in clinical medicine articles.

### SEARCH STRATEGY

The present article followed the search strategy recommended for narrative reviews.<sup>16</sup> Accordingly, PubMed/MEDLINE and Web of Science databases were searched by using the keywords “statistics”, “statistical analysis”, “errors”, “mistakes”, “rheumatology”, “medicine” in combination with appropriate logical connectors. Articles published during the last 5 years until 25<sup>th</sup> January 2019 and those written in the English language were included in this narrative review. The exclusion criteria were as follows: i) letters to editors, ii) commentaries, iii) abstracts and iv) unpublished data. Following the literature search, retrieved articles were assessed for eligibility, and duplicate items or those unrelated to the topic were excluded. Remaining articles’ reference lists were checked, and relevant articles were further included in the narrative review.

### IMPLEMENTING STATISTICAL PERSPECTIVE AT EACH STEP OF A RESEARCH ARTICLE

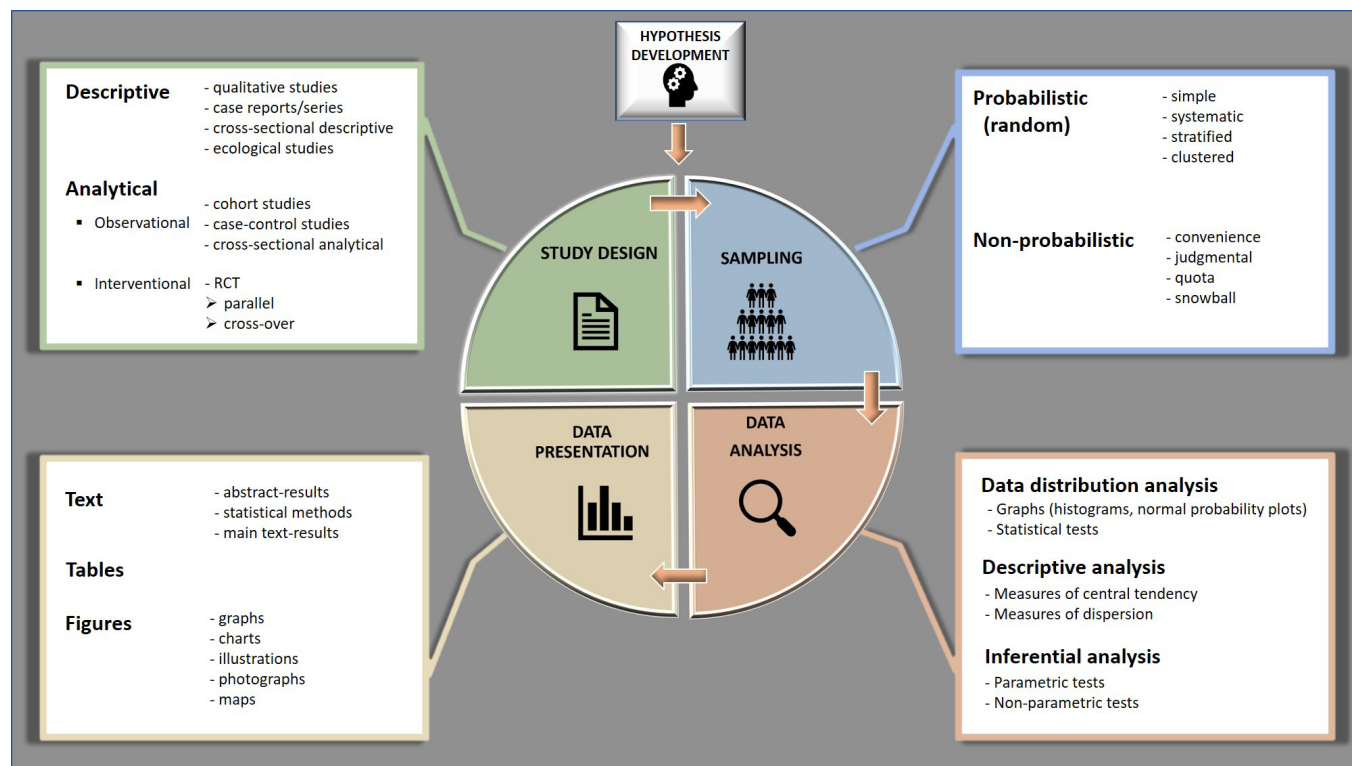
It is of great importance to implement a statistical perspective at each step of any research.<sup>17</sup> These steps can be summarised as research question/hypothesis development, study design, sampling/data collection, data analysis and data interpretation/presentation (*Figures 1 and 2*).

#### Research Question and Hypothesis Development

Every study is designed around a research question. Thereafter, a hypothesis is set upon this question, as the hypothetico-deductive approach is followed by most medical scientific projects.<sup>18</sup> A hypothesis is a testable statement created by the researchers. It is the working instrument of theory and relates theory to observation or vice versa. Hypotheses are relational propositions providing the researcher with a relational statement and a tentative explanation of phenomena.<sup>1</sup>

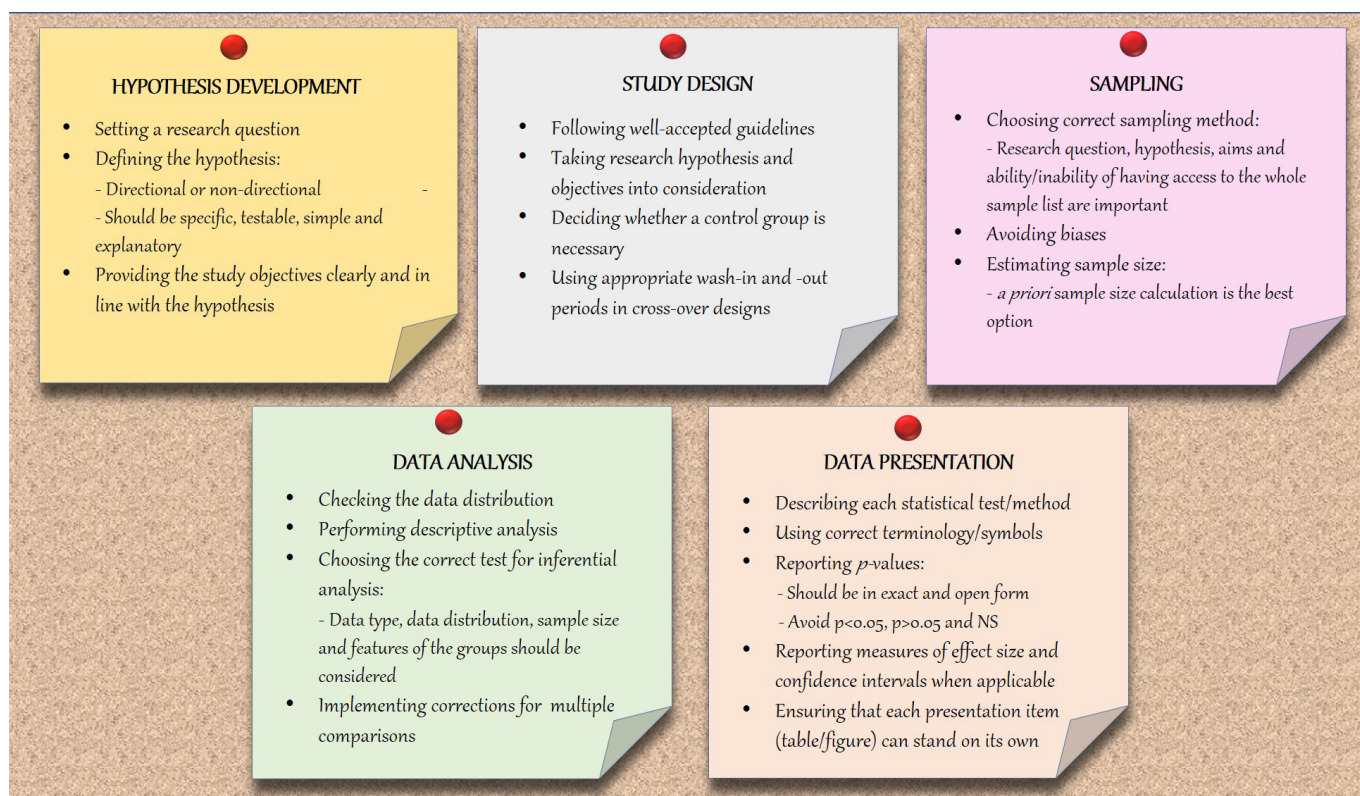
#### What are the features of a good hypothesis?

A good hypothesis must be specific and based on an appropriate research question. It must also be simple, with explanatory power. Most important of all, the developed hypothesis must be testable.<sup>19</sup> After the testing procedure, researchers will either retain the null hypothe-



RCT: Randomised Controlled Trial.

Figure 1. Research steps and statistical considerations.



**Figure 2.** Tips for statistical accuracy in research articles.

sis ( $H_0$  or  $H_N$ ) or reject it in favour of the alternate hypothesis ( $H_1$  or  $H_A$ ).

The null hypothesis states that there is no accurate relation between tested variables. On the other hand, the alternate hypothesis states that  $H_0$  is not correct. Hypothesis testing is necessary to reject the null hypothesis or its alternative. Nevertheless, what should be kept in mind is that not rejecting  $H_0$  does not necessarily mean that the null hypothesis is true, as it might also indicate the insufficiency/deficiency of evidence against the null hypothesis.

The alternate hypothesis can be either directional or non-directional. The latter does not specify the direction of the relationship, whilst the directional hypothesis predicts a direction to the expected findings. At this step, what is already known from the past literature would guide the researchers in deciding the type of the alternative hypothesis.<sup>19</sup>

#### Study Design

Research hypothesis should be tested on an accurate sample, and by using the correct study design. The design of a study can be either descriptive or analytical in nature; either observational or interventional in terms of methodology. Descriptive studies examine the distribution of one

or more variables without providing any causality. These are qualitative studies, case reports/series, cross-sectional descriptive studies (surveys) and ecological studies.<sup>20,21</sup> On the other hand, analytical studies quantify the relationship between variables. Analytical studies can be designed either as observational (non-experimental) or interventional (experimental). Analytical observational studies observe the outcomes of any exposure without the active involvement of the researchers. Cohort studies, analytical cross-sectional studies, and case-control studies are examples of analytic observational research studies.<sup>3,20</sup> On the other hand, interventional studies require the active involvement of the researchers.<sup>20</sup> Randomised controlled trials are the prototypes of interventional studies.<sup>3,20</sup> Two standard types of randomised controlled trials are the parallel and cross-over designed trials.<sup>21</sup> Following well-accepted guidelines such as Consolidated Standards of Reporting Trials (CONSORT), the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE), the Transparent Reporting of Evaluations with Nonrandomised Designs (TREND), and the Statistical Analysis and Methods in the Published Literature (SAMPL) is highly recommended to keep good research practice and standards in researching.<sup>11,22-25</sup>

### Which design is the best for my study?

The research hypothesis and the related aims are major determinants while choosing the design of a study. If researchers target to examine the prevalence/incidence of a certain rheumatic disease or health-related characteristics of patients with this rheumatic disease, a descriptive study design should be set. On the other hand, if the aim is to analyse the effect of a factor (eg, smoking) on the disease-related variables of a rheumatic disease, the research design should be analytic.

Cross-sectional design can be used to determine the frequency of a disease. Moreover, it can also help to analyse the diagnostic procedures or disease-related parameters. However, if the prognostic value of any risk factor will be tested, the best design is the prospective cohort. On some occasions, the time required to observe this prognostic effect is too long. For these instances, retrospective cohort design can be applied. Nevertheless, in retrospectively designed studies, it is hard to eliminate or control the other confounders of disease prognosis.<sup>27</sup> Randomised controlled trials should be the choice if the researchers aim to test the efficacy/effectiveness of any intervention. A parallel or cross-over design can be applied following the randomisation procedure. In a parallel designed trial comparing the efficacy of two biologics (A and B) in patients with ankylosing spondylitis; patients will be assigned either to receive biologic A or biologic B throughout the entire study period. On the other hand, in cross-over designs, patient groups will cross over from biologic A or biologic B during the course of the trial and vice versa. Since every individual serves as his/her control, cross-over design eliminates inter-individual differences and their potential effect on treatment outcomes. However, one important handicap of a cross-over design is the carry-over effect of each intervention. Appropriate use of wash-in and wash-out periods is essential in order to overcome this handicap.<sup>21</sup> Pan et al., in their systematic review, appraised the methodological and reporting quality of randomised controlled trials conducted among patients with rheumatoid arthritis.<sup>28</sup> They reported a high risk of bias and poor adherence to recommended reporting standards for randomised controlled trials.<sup>28</sup> Randomised controlled trials stand as the hallmark of evidence-based medicine and it is of great importance to keep standards in sampling and randomisation while conducting these studies. However, it may be hard to apply randomisation in real-world practice. Therefore, "real" clinical premises and settings are also of value in providing complementary evidence.<sup>29</sup>

### *Sampling and Data Collection*

Sampling is the selection of a sample population (study population) from the whole population (target population). The study sample must represent the whole population properly in order to draw reliable conclusions regarding

the target population by interpreting the results obtained from this sample. Flawed selection procedure causes a lack of representativeness. Therefore, sampling methodology should be set meticulously to achieve an accurate sample population.<sup>30,31</sup>

### Which sampling method should I use for my study?

There are two main methods of sampling: i) probabilistic (random) sampling and ii) non-probabilistic sampling. Probabilistic sampling is the most recommended way of sampling, since this method almost ascertains the representation of the whole population by the sample population. Probabilistic sampling can be further categorised into simple, systematic, stratified and clustered random sampling.<sup>31</sup> Simple random sampling refers to the randomisation of the whole population by using a random numbers table or a computer-based random list. On the other hand, systematic random sampling uses a systematic rule or a fixed interval for randomisation.<sup>30</sup> Using these two sampling methods might have some limitations. If the researchers aim to include subjects from different age groups to their study, simple/systematic random sampling may not fully meet the researchers' needs. In other words, simple sampling may fail to obtain adequate sampling from all age strata in the population. For these instances, stratified random sampling would work much better. Stratified random sampling is a modified simple sampling method, in which the division of the target population into strata is followed by random sampling from each stratum. The researchers divide the whole population into subgroups according to a factor such as age, gender, level of education or diagnosis.<sup>30</sup> Then subjects are selected within each subgroup either through simple or systematic sampling.<sup>31</sup> Research objectives and hypothesis are important while deciding which factor of division will be applied. If the researchers are not able to access the whole sample list due to the large size of the population, cluster sampling should be the way of choice.<sup>30</sup> For example, if a nation-wide research is planned on knee osteoarthritis, the target population can be divided into clusters. Individual centres or general-practitioner practices can be assigned as clusters. Thereafter, other sampling methods are applied in order to select patients with knee osteoarthritis from each cluster/centre.<sup>30</sup> This method can be regarded as an example of multi-stage sampling.<sup>30</sup> If researchers plan to conduct a population-based observational study aiming to determine the risk factors of osteoarthritis, a multi-stage (complex) sampling will be appropriate.

In non-probabilistic sampling, the subjects are selected in a non-systematic process; thus, are usually not representatives of the target population. However, this method is still valuable in answering some research questions or generating new research hypotheses. Non-probabilistic sampling has several types including convenience

(accidental), judgmental (purposive), quota and snow-ball sampling.<sup>30,31</sup> The convenience sampling is the most widely used sampling method in clinical trials. In this method, subjects are selected according to their convenient accessibility. Thus, this method is subject to volunteer bias. Judgmental sampling selects subjects by the choice of the researchers and is subject to investigator bias.<sup>30</sup> In quota sampling, quotas are chosen in order to represent the features of the target population and sampling units are selected to complete the quotas.<sup>31</sup> Snow-ball sampling is used when a target sample cannot be located in a specific place. Each participant is asked to indicate other potential participants.<sup>30,31</sup>

#### How many subjects should I include in my study?

One of the major steps prior to data collection is the sample size estimation. Nour-Eldein reported that almost one-third of the family medicine articles assessed in the study did not mention sample size calculation.<sup>11</sup> Sample size estimation enables the researchers to approximate the enough number of subjects to achieve acceptable study power-thus veracious results as the greater the power the lower the proportion of “false-negatives”.<sup>32,33</sup> Running studies with low statistical power will reduce not only the likelihood of detecting true effects, but also the risk that a nominally statistically significant result (such as  $p < 0.05$ ) actually denotes a ‘real’ true effect.<sup>32</sup> A large enough sample size is required to diminish the type-II error. Type-II error ( $\beta$ -error) refers to the wrong decision when a statistical test fails to reject a false  $H_0$ . The Greek letter  $\beta$  represents the likelihood of making type-II error, that is to say, false-negative rate. For any sample, the maximum amount of false-negative results is desired as being 20%.<sup>3</sup> Since the power of the study is expressed as  $1 - \beta$ , an experiment with a  $\beta$  of 0.20 will have a power of 80%.<sup>34</sup> On the other hand, running studies with more than enough patients would increase the cost and put a needless number of patients at risk of potential harm.<sup>15</sup> Running an *a priori* power analysis by taking significance level, power and effect size into consideration would be the best option to estimate the sample size of the planned study.<sup>32,34</sup>

#### *Data Analysis*

Analysing the data is the most exciting step of research. Using correct statistical methods for data analysis is crucial in order to avoid incoherent, misleading and suboptimal results.

#### What is the way to accurately analyse data?

The first step of accurate data analysis is sound statistical design. If the researchers have performed the above-mentioned procedures including hypothesis development, study design, and sampling properly by taking the study aim and research question into consideration, it means that they are almost halfway through in statis-

tical analysis. There are several statistical packages and numerous statistical tests for data analysis. Researchers should decide which statistical test is appropriate for the data.<sup>35</sup> Bahar et al. evaluated 141 articles published in 6 cytopathology journals in 2014.<sup>2</sup> The most common type of errors in the commission errors (misuse of statistical methods in reporting results) category was the failure to select appropriate statistical methods.<sup>2</sup> Typical examples include using parametric tests for non-parametric data, running independent sample tests for paired data and Pearson’s chi-squared test for inappropriate sample size.<sup>2</sup> Thus, the only possible way to decide the test to use is the recognition of the sample and related data. At that point, researchers should pay attention to data type, data distribution, sample size and dependent/independent groups/variables in the study.<sup>35</sup>

Data type is a major determinant of accurate data analysis. Allam et al. reported that statistical tests are incompatible with the type of data in more than half of the studies (53.2%) evaluated in their study.<sup>35</sup> Variable type can be either qualitative or quantitative. Nominal and ordinal variables are representatives of quantitative data.<sup>36</sup> Variables expressed by nominal or ordinal values are also regarded as categorical variables. For example, in an osteoarthritis study examining the relation of falls with disease severity, occupation is regarded as a nominal variable. On the other hand, Kellgren-Lawrence grade of osteoarthritis is an ordinal variable. Quantitative variables are numeric and subdivided to continuous and discrete variables. Discrete items are expressed by whole numbers, whereas continuous items can be recorded as any value.<sup>36</sup> Regarding the above-mentioned study, the number of falls is discrete data, whilst body weight or height are continuous variables. The dependency/independency of each variable or tested groups is another important point while choosing the statistical test. The decision of running paired or unpaired tests are made according to this point. On the other hand, the decision of performing parametric or non-parametric tests is based upon the distribution of data.

#### Which statistical tests should I apply to my data?

Descriptive statistics is the first step in statistical analysis. It provides description about the measures of central tendency (mean, median, mode), as well as the measures of dispersion (standard deviation, variance, standard error of mean, range, percentiles).<sup>36</sup> Mean is the arithmetic average of a certain variable and may be affected by the extreme variables/outliers. Median refers to the 50<sup>th</sup> percentile, while the inter-percentile (inter-quartile) range represents the observations between the 75<sup>th</sup> and 25<sup>th</sup> percentile (the middle 50%) of the observations.<sup>36</sup> Data distribution should be tested in order to decide which of the above-mentioned measures will be used to report the data.<sup>14</sup> Testing data distribution is also re-

quired before moving to the second step—the inferential analysis. Data distribution is evaluated by statistical tests (ie, Kolmogorov-Smirnov and Shapiro-Wilks tests) and graphical analyses including histograms and normal probability plots. Histogram is a graphical technique on which the skewness and kurtosis of a data set can be evaluated. The standard normal distribution curve has a symmetrical bell shape.<sup>3,14,36</sup> Skewness is the degree of distortion from this bell-shaped curve and regarded as the measure of asymmetry.<sup>36</sup> Kurtosis is the degree of tailedness of a distribution. In addition to the graphical evaluation, it is of value to assess skewness and kurtosis values by using representative indexes as well.

Inferential statistics uses data of a sample population to make inferences in the larger collection of the population, and can be applied after testing the distribution of data. Parametric tests are applied for normally distributed quantitative data, whilst nonparametric tests are performed when the assumption of normality is not met.<sup>36</sup> One of the most common misconceptions in data analysis is to regard nonparametric statistical tests as distribution/assumption-free or less powerful than the parametric tests. Instead, large-sample approximate tests are assumed to follow known distributions and, on some occasions, nonparametric tests can be more powerful than their parametric counterparts.<sup>7</sup>

If a comparative analysis is the case, it is important whether the groups are dependent or not.<sup>36,37</sup> Parametric tests used to compare two paired groups and more than 2 matched groups are paired *t*-test and repeated measures analysis of variance (ANOVA), respectively. Nonparametric counterparts of these tests are Wilcoxon's signed rank test and Friedman's test. On the other hand, parametric tests used to compare two unpaired groups and more than 2 unmatched groups are unpaired *t*-test and ANOVA, respectively. Nonparametric alternatives of these tests are the Mann-Whitney *U* test and Kruskal-Wallis test.<sup>2,36,38</sup> In case of multiple inferences on the same data sets or for comparisons of more than 2 groups, a post-hoc correction should be applied,<sup>35</sup> since testing multiple hypotheses lead an inflated type-I error rate (false positivity), unless correction methods are used.<sup>7</sup>

Categorical data presented in frequencies/percentages are tested by using the Pearson's chi-squared test, Fisher's exact or McNemar's tests, etc.<sup>36</sup> If more than 20% of the cells have frequencies <5 (which indicates at least 5 observations in each category), Fisher's exact test should be used instead of the Pearson's chi-squared test.<sup>5,11</sup> As for correlation studies, normally distributed data is tested by Pearson's correlation analysis, whereas the Spearman's correlation test is used for data beyond normal distribution.<sup>2,36</sup>

#### *Data Interpretation and Presentation*

Data presentation is as important as data analysis in research. Providing the readers with proper presentations would allow accurate transmission of scientific evidence among researchers. Gunel Karadeniz et al. founded that 147/157 articles published in radiology journals had at least one statistical error, and statistical error distribution did not differ between journals with impact factors  $\geq 2$  and  $< 2$ .<sup>9</sup> The most common type of error was incorrect data summarising (65.6%).<sup>9</sup> In order to summarise the data correctly, researchers should understand and interpret the results accurately. On several occasions, *p*-values, which are definitely the most common statistical measure reported in articles are misinterpreted by the authors.<sup>7,12,39,40</sup>

#### How should I interpret *p*-value results?

Significance is often discussed in terms of *p*-values (ranging from 0 to 1). On the basis of a prespecified significance level ( $\alpha$ ), *p*-values less than  $\alpha$  indicates statistically significance in which the null hypothesis is rejected.<sup>7</sup> The smaller the *p*-value, the greater the incompatibility of the data with  $H_0$ .<sup>41</sup> However, it does not imply that the difference is big or relevant.<sup>42</sup> On the other hand, larger *p*-values suggest that there is insufficient evidence/information to reject  $H_0$ . At that point, it is a mistake to focus solely on the results of null-hypothesis significance testing, that is to say, the *p*-value.<sup>7,25,41</sup> Researchers should also pay attention to the magnitude of association—usually referred to as the effect size. The American Statistical Association's statement on statistical significance and *p*-values indicates that a *p*-value, by itself, does not measure the size of effect or the importance of a finding.<sup>7,41</sup> For a better interpretation of the results, it is of value to take measures of obtained effect sizes, minimal clinically important differences and confidence intervals into consideration, instead of relying solely on the *p*-values.<sup>7,38,43</sup>

#### How should I present my statistical output?

Satisfactory presentation of statistical output is an issue of concern for researchers. There are several ways for data presentation including texts, tables, and figures (graphs, charts, illustrations, photographs, and maps). Essentially, authors start presenting the study data as a text inside the abstract of their paper. At that point, being concise and explanatory is the key point. Firstly, a brief presentation of the sampling results such as sample size, age, gender, disease duration should be provided. Thereafter, the main finding of the study should be given. Results with regard to the secondary objective(s) may follow the main finding. Wherever possible/applicable, findings should be provided along with relevant *p*-values, correlation coefficients and confidence intervals inside the abstract.<sup>44</sup>

The second step should be the presentation of statistical methodology. Statistical methodology should be given at the end of the Methods section as a separate paragraph.<sup>38</sup> Each statistical methods paragraph should be unique, as one “generic” paragraph would not fit all studies.<sup>45</sup> Gunel Karadeniz et al. showed that almost 10% of the articles failed to provide statistical methodology.<sup>9</sup> In this section of the article, data analysis technique(s), statistical software package used, accepted significance level should be provided. When multiple statistical tests or techniques are applied, the authors should state which test is used for which data, instead of using phrases such as “where applicable” or “where appropriate”.<sup>3</sup> When multiple comparisons are performed, a multiple comparison correction/a level correction should be provided.<sup>35</sup>

The main arena for data presentation is the Results section of the paper. The Results text should be written after interpreting the data correctly, and by using appropriate terminology. Unfortunately, misinterpretation and misuse of statistical terminology/symbols are common in scientific articles.<sup>35</sup> It is an issue of concern to focus on reporting positive statistical results belonging to secondary/tertiary aims in order to motivate the paper. However, regardless of the results being positive or negative, the greater focus should be given on the data related to the primary hypothesis.<sup>45</sup> Researchers should also avoid selective reporting practices, as non-reported secondary findings and/or non-significant results can also be important and informative for readers, future researchers, overall for community and science.<sup>6</sup>

The results section should start with summarising each variable by descriptive statistics. Normally distributed variables are presented with mean and standard deviation, whereas skewed data are presented in median and range/interquartile range. It is of value to provide not only the size of the range but also the minimum and maximum values of ranges or upper and lower boundaries of interquartile ranges.<sup>25</sup> The second step is the presentation of inferential statistical test results. At this step, authors should pay attention to present the results as much detailed as possible. For instance, instead of stating as “Minimum joint space width of the medial tibiofemoral joint differed significantly between patients with knee osteoarthritis and controls”, results should be expressed as “Minimum joint space width of the medial tibiofemoral joint was significantly narrower in patients with knee osteoarthritis than that in controls ( $p=0.02$ ).<sup>44</sup> Total sample and group sizes for each analysis should be provided.<sup>25</sup> A common mistake while interpreting/presenting the results of randomised controlled trials or interventional studies is the use of within-group differences to infer between-group differences. For example, researchers may intend to evaluate the effect of aerobic training in fibromyalgia patients via a randomised controlled trial. The

results may reveal that the experimental group showed significant improvement in outcome measures, while the controls showed no significant improvement over time. By solely using these within-group data, it will be a major mistake to conclude that aerobic training is effective in fibromyalgia.<sup>45</sup> Rather, researchers should base their inference on the significance of the difference between groups.<sup>7</sup> Gelman and Stern highlighted this important point in their article entitled “The difference between ‘significant’ and ‘not significant’ is not itself statistically significant”.<sup>46</sup> Researchers should also avoid using the word “correlated” when only a comparative analysis was carried out without any correlation analysis.<sup>2</sup>

As for the interpretation of correlation analysis, researchers should not miss that a correlation itself does not necessarily represent a causal relationship; instead, it is regarded as one of the necessary conditions of causation.<sup>10,42</sup> Therefore, researchers should avoid the word “predicted” when only a correlation analysis was performed.<sup>2</sup> While reporting coefficients in association with correlation and regression, or reporting effect size measures such as absolute/relative risks, incidence/survival rates, odds/hazard ratios, confidence intervals (measure of precision) should be provided.<sup>2,25</sup> Regression equation should be given for simple or multiple regression analysis.<sup>25</sup> For multiple regression, variable selection process (eg, forward-stepwise) should be reported.<sup>11,25</sup> As for diagnostic accuracy studies, the quantification of the statistical tests can be obtained by sensitivity, specificity, positive/negative predictive values, positive/negative likelihood ratios, odds ratio and/or the area under the receiver operating curve (ROC).

A big amount of errors in data presentation is related to  $p$ -values. Regarding the articles published in radiology journals, common mistakes related to  $p$ -values are: lack of reporting  $p$ -values, giving  $p$ -values in closed forms (ie,  $p<0.05$  or  $p>0.05$ ), incorrect expression of  $p$ -values (ie,  $p<0.0005$  or  $p=0.000$ ), providing  $p$ -values at conclusion section of the manuscript and using colon instead of equal sign in presentation ( $p:0.001$  instead of  $p=0.001$ ). It is recommended to describe  $p$ -values to 3 decimal places and exact values.<sup>10</sup> On the other hand, it is of great value to provide the readers with the size of the effects studied. Reporting effect sizes or mean differences with confidence intervals would allow the readers to evaluate the power of the study and to clarify whether enough subjects were tested.<sup>6,7</sup>

Creating tables is an appropriate way to present large amounts of data. On the other hand, figure presentations such as graphs and charts are good at depicting the relation between series of numbers. Scatter plots are used to depict the association between two variables. Bar graphs are appropriate to indicate comparative results of different groups. They can be created horizontally or vertically, where the length (height) of the bar represents



the amount of information. Pie charts are useful for representing data classified in different categories. Line plots are used to represent time-series data. Box and whisker charts are useful for presenting nonparametric data.<sup>47</sup> Each table or graph should be self-explanatory and stand on its own regardless of the relevant text. Every presenting unit (graph, chart, table, figure) should have an accurate title. All abbreviations should be given in full at footnotes or figure legends. Authors should pay attention to referring each table or figure in the relevant part of the text.

## CONCLUSION

The guidance of statistics is required at each step of research. Implementing statistics would ensure adequate data collection, appropriate statistical analysis, and presentation. Statistical misuse is common among practices of medicine. Further research is required to better identify the misuse of statistical methods in rheumatology papers. It is of utmost importance for rheumatologists to be aware of the common errors in statistical use. Moreover, basic knowledge of statistical tools/techniques is essential. Supervision by an expert on biostatistics is recommended from the beginning of each research until the end of the whole process. In order to increase the statistical accuracy in scientific articles, it is useful to run a statistical review and/or provide the assigned reviewers with statistical checklists. Most importantly, basic education on statistical use and research methodology should be implemented in both under-graduate and post-graduate education in medicine.

## CONFLICT OF INTEREST

The author declares no conflict of interest.

## REFERENCES

1. Brown AW, Kaiser KA, Allison DB. Issues with data and analyses: Errors, underlying themes, and potential solutions. *Proc Natl Acad Sci U S A* 2018;115:2563-70. [https://doi.org/10.1073/pnas.1708279115] [PMID: 29531079] [PMCID: PMC5856502]
2. Bahar B, Pambuccian SE, Barkan GA, Akdas Y. The Use and Misuse of Statistical Methods in Cytopathology Studies: Review of 6 Journals. *Lab Med* 2019;50:8-15. [https://doi.org/10.1093/labmed/lmy036] [PMID: 30020520]
3. Bajwa SJ. Basics, common errors and essentials of statistical tools and techniques in anesthesiology research. *J Anaesthesiol Clin Pharmacol* 2015;31:547-53. [https://doi.org/10.4103/0970-9185.169087] [PMID: 26702217] [PMCID: PMC4676249]
4. Carmona-Bayonas A, Jimenez-Fonseca P, Fernández-Somoano A, Álvarez-Manceño F, Castañón E, Custodio A, et al. Top ten errors of statistical analysis in observational studies for cancer research. *Clin Transl Oncol* 2018;20:954-65. [https://doi.org/10.1007/s12094-017-1817-9] [PMID: 29218627]
5. Choi SW, Cheung CW. Don't judge a book by its cover, don't judge a study by its abstract. Common statistical errors seen in medical papers. *Anaesthesia* 2016;71:843-5. [https://doi.org/10.1111/anae.13506] [PMID: 27156782]
6. Hupé JM. Statistical inferences under the Null hypothesis: common mistakes and pitfalls in neuroimaging studies. *Front Neurosci* 2015;9:18. [https://doi.org/10.3389/fnins.2015.00018] [PMID: 25745383] [PMCID: PMC4333770]
7. George BJ, Beasley TM, Brown AW, Dawson J, Dimova R, Divers J, et al. Common scientific and statistical errors in obesity research. *Obesity (Silver Spring)* 2016;24:71-90. [https://doi.org/10.1002/oby.21449] [PMID: 27028280] [PMCID: PMC4817356]
8. Gosselin RD. Guidelines on statistics for researchers using laboratory animals: the essentials. *Lab Anim* 2019;53:28-42. [https://doi.org/10.1177/2F0023677218783223]
9. Gunel Karadeniz P, Uzabacı E, Atış Kuyuk S, Kaskır Kesin F, Can FE, Seçil M, et al. Statistical errors in articles published in radiology journals. *Diagn Interv Radiol* 2019;25:102-8. [https://doi.org/10.5152/dir.2018.18148] [PMID: 30582574] [PMCID: PMC6411271]
10. Lee S. Avoiding negative reviewer comments: common statistical errors in anesthesia journals. *Korean J Anesthesiol* 2016;69:219-26. [https://doi.org/10.4097/kjae.2016.69.3.219] [PMID: 27274365] [PMCID: PMC4891532]
11. Nour-Eldein H. Statistical methods and errors in family medicine articles between 2010 and 2014-Suez Canal University, Egypt: A cross-sectional study. *J Family Med Prim Care* 2016;5:24-33. [https://doi.org/10.4103/2249-4863.184619] [PMID: 27453839] [PMCID: PMC4943144]
12. Turkiewicz A, Luta G, Hughes HV, Ranstam J. Statistical mistakes and how to avoid them - lessons learned from the reproducibility crisis. *Osteoarthritis Cartilage* 2018;26:1409-11. [https://doi.org/10.1016/j.joca.2018.07.017] [PMID: 30096356]
13. Arifin WN, Sarimah A, Norsaadah B, Najib Majidi Y, Siti-Azrin AH, Kamarul Imran M, et al. Reporting Statistical Results in Medical Journals. *Malays J Med Sci* 2016;23:1-7. [https://doi.org/10.21315/mjms2016.23.5.1] [PMID: 27904419] [PMCID: PMC5101968]
14. Bianchi MT, Phillips AJ, Wang W, Klerman EB. Statistics for Sleep and Biological Rhythms Research. *J Biol Rhythms* 2017;32:7-17. [https://doi.org/10.1177/0748730416670050] [PMID: 27836938] [PMCID: PMC5501416]
15. Celik S, Yazici Y, Yazici H. Are sample sizes of randomised clinical trials in rheumatoid arthritis too large? *Eur J Clin Invest* 2014;44:1034-44. [https://doi.org/10.1111/eci.12337] [PMID: 25207845]
16. Gasparyan AY, Ayyavazyan L, Blackmore H, Kitas GD. Writing a narrative biomedical review: considerations for authors, peer reviewers, and editors. *Rheumatol Int* 2011;31:1409-17. [https://doi.org/10.1007/s00296-011-1999-3] [PMID: 21800117]
17. Coskun Benlidayi I. Implement statistics at each step of your research. *Rheumatol Int* 2019 Jul;39(7):1303-4. [https://doi.org/10.1007/s00296-019-04327-3] [PMID: 31111292]
18. Tariq MU. Hypothetico-deductive method: A comparative analysis. *JOBARI* 2015;7:228-31.
19. Mourougan S, Sethuraman K. Hypothesis development and testing. *IOSR-JBM* 2017;19:34-40. [https://doi.org/10.9790/487X-1905013440]
20. The Centre for Evidence-Based Medicine (CEBM) -Study Designs. Available from: <https://www.cebm.net/2014/04/study-designs/> [Last accessed on 2019 May 26].
21. Spieth PM, Kubasch AS, Penzlin AI, Illigens BM, Barlinn K, Siepmann T. Randomised controlled trials- a matter of design. *Neuropsychiatr Dis Treat* 2016;12:1341-9. [https://doi.org/10.2147/NDT.S101938] [PMID: 27354804] [PMCID: PMC4910682]
22. Schulz KF, Altman DG, Moher D; CONSORT Group. Consort 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Ann Intern Med* 2010;152:726-32. [https://doi.org/10.7326/0003-4819-152-11-201006010-00232] [PMID: 20335313]
23. Fuller T, Peters J, Pearson M, Anderson R. Impact of the transparent reporting of evaluations with nonrandomised designs reporting guideline: Ten years on. *Am J Public Health* 2014;104:e110-7. [https://doi.org/10.2105/AJPH.2014.302195] [PMID: 25211744] [PMCID: PMC4202955]
24. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The strengthening the report-

- ing of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Ann Intern Med* 2007;147:573-7. [https://doi.org/10.7326/0003-4819-147-8-200710160-00010] [PMID: 17938396]
25. Lang T, Altman D. Statistical analyses and methods in the published literature: The SAMPL guidelines. *Medical Writing* 2016;25:31-6. [https://doi.org/10.18243/eon/2016.9.7.4]
  26. The EQUATOR Network | Enhancing the Quality and Transparency of Health Research. Available from: <http://www.equator-network.org> [Last accessed on 2019 May 25].
  27. Sut N. Study designs in medicine. *Balkan Med J* 2014;31(4):273-7. [https://doi.org/10.5152/balkanmedj.2014.1408] [PMID: 25667779] [PMCID: PMC4318396]
  28. Pan X, Lopez-Olivo MA, Song J, Pratt G, Suarez-Almazor ME. Systematic review of the methodological quality of controlled trials evaluating Chinese herbal medicine in patients with rheumatoid arthritis. *BMJ Open* 2017;7:e013242. [https://doi.org/10.1136/bmjopen-2016-013242] [PMID: 28249848] [PMCID: PMC5353312]
  29. Coskun Benlidayi I. Reliability of real-world data. *Rheumatol Int* 2019;39:583-4. [https://doi.org/10.1007/s00296-018-4230-y] [PMID: 30604203]
  30. Efil M, Negida A. Sampling methods in Clinical Research; an Educational Review. *Emerg (Tehran)* 2017;5:e52. [PMID: 28286859] [PMCID: PMC5325924]
  31. Martínez-Mesa J, González-Chica DA, Duquia RP, Bonamico RR, Bastos JL. Sampling: how to select participants in my research study? *An Bras Dermatol* 2016;91:326-30. [https://doi.org/10.1590/abd1806-4841.20165254] [PMID: 27438200] [PMCID: PMC4938277]
  32. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013;14:365-76. [https://doi.org/10.1038/nrn3475] [PMID: 23571845]
  33. Burt T, Button KS, Thom H, Noveck RJ, Munafò MR. The Burden of the "False-Negatives" in Clinical Development: Analyses of Current and Alternative Scenarios and Corrective Measures. *Clin Transl Sci* 2017;10:470-9. [https://doi.org/10.1111/cts.12478] [PMID: 28675646] [PMCID: PMC6402187]
  34. Beck TW. The importance of a priori sample size estimation in strength and conditioning research. *J Strength Cond Res* 2013;27:2323-37. [https://doi.org/10.1519/JSC.0b013e318278eea0] [PMID: 23880657]
  35. Allam RM, Noaman MK, Moneer MM, Elattar IA. Assessment of Statistical Methodologies and Pitfalls of Dissertations Carried Out at National Cancer Institute, Cairo University. *Asian Pac J Cancer Prev* 2017;18:231-7. [https://doi.org/10.22034/APJCP.2017.18.1.231] [PMID: 28240524] [PMCID: PMC5563106]
  36. Ali Z, Bhaskar SB. Basic statistical tools in research and data analysis. *Indian J Anaesth* 2016;60:662-9.
  37. Weissgerber TL, Garovic VD, Milin-Lazovic JS, Winham SJ, Obradovic Z, Trzeciakowski JP, et al. Reinventing Biostatistics Education for Basic Scientists. *PLoS Biol* 2016;14:e1002430. [https://doi.org/10.1371/journal.pbio.1002430] [PMID: 27058055] [PMCID: PMC4825954]
  38. Habibzadeh F. Statistical Data Editing in Scientific Articles. *J Korean Med Sci* 2017;32:1072-6. [https://doi.org/10.3346/jkms.2017.32.7.1072] [PMID: 28581261] [PMCID: PMC5461308]
  39. Dvir Z. Difference, significant difference and clinically meaningful difference: The meaning of change in rehabilitation. *J Exerc Rehabil* 2015;11:67-73. [https://doi.org/10.12965/jer.150199] [PMID: 25960978] [PMCID: PMC4415752]
  40. Wilkinson M. Distinguishing between statistical significance and practical/clinical meaningfulness using statistical inference. *Sports Med* 2014;44:295-301. [https://doi.org/10.1007/s40279-013-0125-y] [PMID: 24248505]
  41. Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. *Am Stat* 2016;70:129-33. [https://doi.org/10.1080/00031305.2016.1154108]
  42. Schwen LO, Rueschenbaum S. Ten quick tips for getting the most scientific value out of numerical data. *PLoS Comput Biol* 2018;14:e1006141. [https://doi.org/10.1371/journal.pcbi.1006141] [PMID: 30307934] [PMCID: PMC6181270]
  43. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol* 2017;82:128-36. [https://doi.org/10.1016/j.jclinepi.2016.11.016] [PMID: 27986610]
  44. Andrade C. How to write a good abstract for a scientific paper or conference presentation. *Indian J Psychiatry* 2011;53:172-5. [https://doi.org/10.4103/0019-5545.82558] [PMID: 21772657] [PMCID: PMC3136027]
  45. Zinsmeister AR, Connor JT. Ten common statistical errors and how to avoid them. *Am J Gastroenterol* 2008;103:262-6. [https://doi.org/10.1111/j.1572-0241.2007.01590.x] [PMID: 18289193]
  46. Gelman A, Stern H. The difference between "significant" and "not significant" is not itself statistically significant. *Am Stat* 2006;60:328-31. [https://doi.org/10.1198/000313006X152649]
  47. In J, Lee S. Statistical data presentation. *Korean J Anesthesiol* 2017;70:267-6. [https://doi.org/10.4097/kjae.2017.70.3.267] [PMID: 28580077] [PMCID: PMC5453888]