Statistical errors in articles published in radiology journals

Pinar Günel Karadeniz ©
Ender Uzabacı ©
Sema Atış Kuyuk ©
Fisun Kaskır Kesin ©
Fatma Ezgi Can ©
Mustafa Seçil ©
İlker Ercan ©

From the Department of Biostatistics (P.G.K. gunelpinar@yahoo.com), SANKO University School of Medicine, Gaziantep, Turkey; Department of Biostatistics (E.U.), Uludağ University School of Veterinary Medicine, Bursa, Turkey; Department of Biostatistics (S.A.K., F.K.K., F.E.C.) Uludağ University, Health Science Institute, Bursa, Turkey; Department of Biostatistics (E.E.) Uludağ University School of Medicine, Bursa, Turkey.

Received 29 March 2018; revision requested 10 April 2018; last revision received 28 May 2018; accepted 12 June 2018.

Published online 17 December 2018.
DOI 10.5152/dir.2018.18148


Statistical science is utilized in all processes of scientific studies from the planning to the reporting stages. The fact that statistics is essential in scientific studies today indicates that the correct use of statistical procedures is also highly important. Medical authorities also emphasize the importance of statistics and state that physicians should at least be good readers of statistics. Researchers publishing in scientific journals, who do not have sufficient knowledge of statistics, may make mistakes in their use of statistical science at any step, including planning, design, execution, analysis, and presentation of data. Although mostly unrecognizable by the readers following the literature, the vast majority of the articles in the medical literature contain statistical errors and omissions. Some of these errors directly affect the results, while others are presentation errors in the representation or terminology, not having a major influence on the result (1). In both situations, these mistakes should be avoided.

Since the 1960s and 1970s, many researchers, wishing to draw attention to the errors and omissions in statistics and methodology, have investigated the most commonly used statistical methods in medical journals and emphasized the importance of correct use of statistics in scientific publications, and have published their research and proposals on this topic. Some
Some authors emphasized in their studies the importance of consulting a statistician, journal editors, and those involved in scientific publications by enabling scientists, statisticians, and journal editors to directly use the results of these studies in their clinical practice. In this context, both scientific journals and researchers who carry out the task of transmission of scientific knowledge carry a great responsibility to avoid mistakes.

The aim of this study is to evaluate the articles in radiology journals indexed in the Science Citation Index Expanded (SCI-E) in terms of statistical errors. Thus, we aim to contribute to the production of high quality scientific publications by enabling scientists, journal editors, and those involved in the article evaluation process in radiology journals to be sensitive and careful about statistical errors.

Methods

When examining the literature related to statistical errors in scientific studies in medicine, it was found that about half of the articles contained statistical errors. McGuigan (14) reported that 40% of 164 papers in the British Journal of Psychiatry contained statistical errors. Glantz (4) showed that the error rate in the articles that used statistical techniques in Circulation Research and Circulation was about 50% (61% and 44%). Gore et al. (3) in their study of critical assessment of articles in British Medical Journal from January to March 1976 reported that 52% of 62 papers included at least one statistical error. Lukiae and Maruśiæ (27) found out that statistics were not satisfactory in 63% of 144 articles published in the Croatian Medical Journal. Simundic and Nicolac (10) reported that at least one error was observed in 48 of 55 (87%) manuscripts submitted to the Biochimia Medica Journal. Erkan et al. (11) reported that statistical errors were found in 173 of the 181 manuscripts submitted to Turkish Clinics Journal of Medical Sciences 96%. The median error rate was 0.58 according to these reference studies. In the light of this information, it was decided that the number of articles to be examined should be 158 (n=2*pqd), for the sample size in our study at α = 0.05 significance level and d = 0.077 margin of error with reference to the P = 0.57 rate (28).

The Thomson Reuters Clarivate Analytics database includes 20 radiology journals indexed in SCI or SCI-E, with the word “radiology” in the journal title. In this study, a total of 157 articles were reviewed randomly in these 20 journals, which comprised four articles per year from the articles published in 2016–2017. However, there was only one research article published in a journal in 2017, so just one article for this journal was reviewed for that year.

Author surname was used for randomization in article selection. The article selection algorithm is as follows. Step I: a random article was selected from the first issue of the year 2016 of the first journal in the alphabetical order. Step II: the first letter of the surname of the first author of the first selected article was used to determine the next article to be selected; this was determined as the “random letter”. Step III: the subsequent article in which the first letter of the first author’s name was the “random letter” was selected for reviewing. Step IV: in the last article the first letter of the first author’s surname was again determined as the new “random letter”. Thereafter, Steps III, IV, and V were repeated until the sample size determined in the sampling process was reached. Sampling was done in such a way that an equal number of articles were taken from issues every year, taking into account the number of the articles in a year.

Selected articles were evaluated jointly by five researchers who are biostatistics experts (P. Gunes Karadeniz, E. Uzabaci, S. Atis Kuyuk, F. Kaskir Kesin, F.E. Can) for statistical errors regarding P values and statistical tests, and errors in terminology and other errors related to interpretation. The articles were first shared by the individual researchers then evaluated by the five researchers as a group. Classification of statistical errors in articles was done in line with previous studies by Erkan et al. (1, 11–13). The statistical errors identified by each researcher were confirmed by the entire study team. At this point, it can be said that there was full agreement among the researchers.
Statistical errors in radiology journals

Therefore, there was no need to calculate inter-rater reliability.

Statistical errors are classified as below:

Errors related to P values: P values given in closed form (e.g., $P < 0.01$, $P < 0.05$, $P > 0.05$), non-reported $P$ values, incorrect $P$ values, and incorrect representation of $P$ values (e.g., $P = 0.000, P < 0.0005$).

Errors related to tests: undefined statistical test, incorrect name of the statistical test, statistical technique defined but not used, use of incorrect test, and statistical analysis required but not performed.

Other errors: mathematical notation errors (e.g., using “,” instead of “.” as a decimal point, using “:” instead of “=” while representing sample size or $P$ value $n=120$ or $P<0.0002$), statistical symbol errors (e.g., using $X^2$ instead of $\chi^2$ while showing chi-square test statistics), incomprehensible statistical terms (e.g., presentation of descriptive statistics without explaining which statistics they are; mean±standard deviation or mean±standard error), inappropriate interpretation (e.g., stating there is correlation between two variables when $P > 0.05$), errors in statistical terminology (e.g., stating that “Pearson test was used for measuring correlation”), errors in summarizing data (e.g., when using a parametric test, it is common to incorrectly give median and min-max values instead of mean and standard deviation as descriptive statistics or, conversely, when a nonparametric test is used, it is common to incorrectly give mean and standard deviation instead of median and min-max values), and presentation of the statistical method, analysis and results in the incorrect section of the manuscript (e.g., giving $P$ values at discussion or conclusion parts of the manuscripts).

Statistical error rates were obtained by taking all the articles assessed into account. In addition, in order to examine whether the error rates of the articles published in the radiology journals differed according to the Impact Factor (IF), the journals from which the articles were taken were divided into two groups, namely journals with IF $\geq 2$ and journals with IF $<2$. The error rates of these groups were compared with chi-square and Fisher's exact test.

Results

Of the 157 articles published in radiology journals, there was at least one statistical error in 147. The most frequently encountered error was errors in summarizing data, with a rate of 66% ($n=103$). This was followed by incorrect representation of $P$ values with a rate of 42% ($n=66$). The least frequently encountered error was statistical symbol errors with a rate of 3% ($n=5$).

The results of statistical comparisons made on the basis of the statistical error distributions in the articles published in radiology journals with IF $\geq2$ and IF $<2$ are given in Table 1. There was no statistically significant difference between the groups with IF $\geq2$ and IF $<2$ in regards to statistical errors. Statistical error distributions in similar studies are given in Table 2. Statistical error rates in radiology journals are remarkable especially in representing and reporting the $P$ values, in reporting the name of the statistical test, in summarizing data and in statistical terminology.

Discussion

In this study, statistical errors in articles published in radiology journals indexed in SCI and SCI-E were examined. The accuracy and reliability of published scientific studies is very important for scientists who will make use of the results of these studies. Therefore, published scientific studies should be screened for statistical errors and necessary care should be given to statistics. As Bland argued, “bad statistics leads to bad research and bad research is unethical.” Poor scientific studies should be prevented from turning into bad medicine and accurate research in evidence-based medical practice should be increased (29).

Many studies have been published evaluating the statistical procedures used in scientific articles. When the studies assessing the statistical errors are considered, it may be seen that some of them investigated the errors made in publications in general medicine and some investigated the errors made in articles published in journals dealing with a certain branch of medicine. In this study, statistical errors in publications in the field of radiology were examined.

In the articles we reviewed in radiology journals, the most frequently encountered errors were made in summarizing data with a rate of 65.61%. Previous studies have also reported that the errors in summarizing data are the most frequent errors, with a 28.11% rate in general medicine journals and a 57.84% rate in journals in veterinary science (12, 13). Hanif and Ajmal (30) showed the rate of inadequate and inaccurate presentation of descriptive statistics as 16.25% in their work in local medical journals in Pakistan.

In radiology, diagnoses are usually based on quantitative data. In their study, Medina and Zurakowski (31) emphasized that standard error of mean was used incorrectly instead of standard deviation when summarizing data to make the variability of the data look tighter. In addition, when using a parametric test, it is common to incorrectly give median and min-max values instead of mean and standard deviation as descriptive statistics or, conversely, when a nonparametric test is used, it is common to incorrectly give mean and standard deviation instead of median and min-max values. Therefore, it is important to understand the correct use of basic statistics in order to avoid errors in summarizing data. A well-designed, well-executed scientific study deserves a good presentation. No matter how well you execute your study, it will lose importance if the results are not analyzed or presented correctly (15).

When errors related to $P$ value were considered, the most frequent error was incorrect representation of $P$ values with a rate of 42.04%. Ercan et al. (12, 13) reported this rate as 18.43% in medical journals and 37.25% in veterinary journals. Incorrect representation of $P$ values is a problem that leads to reduced confidence in the study. The second most frequent error was nonreported $P$ values with a rate of 25.48%. This error had rates of 22.12% and 44.12% in the studies of Ercan et al. (12, 13) in medical journals and in veterinary journals, respectively. In these cases, suspicion of the inaccuracy of statistical tests applied in studies arises. In their studies this ratio was reported as 13.36% and 8.82% in medical journals and veterinary journals, respectively, and it was emphasized that because this ratio was only obtained from the articles where $P$ value could be checked, this ratio may actually be even higher (12, 13). This is similar for the articles we examined in this study. $P$ values were given in closed form in 16.56% of the articles that we reviewed. This error was encountered in 15.21% of the articles in medical journals and 49.02% of the articles in veterinary journals examined by Ercan et al. (12, 13). Hanif and Ajmal (30) reported this error in local medical journals in Pakistan as 16.25%, while McGuigan (14) reported it as 51.22% in the study on the articles in the British Journal of Psychiatry. Reporting $P$ values in closed form causes the
Errors in summarizing data | 40 (58.0) | 63 (71.6) | 0.075 | 103 (65.6) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomprehensible statistical terms</td>
<td>9 (13.0)</td>
<td>21 (23.9)</td>
<td>0.132</td>
<td>30 (19.1)</td>
</tr>
<tr>
<td>Errors in statistical terminology</td>
<td>10 (14.5)</td>
<td>20 (22.7)</td>
<td>0.272</td>
<td>30 (19.1)</td>
</tr>
</tbody>
</table>

The rate of this kind of error was 12.10% in our study. In the studies of Ercan et al. (12, 13) in medical and veterinary journals it was found as 3.23% and 9.31%, respectively, while in Hanif and Ajmal’s study it was 12.50% (30). In 9.55% of the articles we examined, the statistical technique was used but not defined. Ercan et al. (12, 13) reported this rate as 11.52% in medical journals and 15.69% in veterinary journals. Hanif and Ajmal (30) found the rate of this error as 26.25%. In 5.73% of the articles we examined, the statistical technique was defined but not used. Frequency of this error was 2.30% in medical journals and 3.43% in veterinary journals in the studies of Ercan et al. (12, 13), while it was 21.25% in Hanif and Ajmal’s study (30). In 8.28% of the studies we examined, a statistical analysis was required but not performed. Ercan et al. (12, 13) found the rate of this kind of error as 17.51% in medical journals and 1.96% in veterinary journals. There is no scientific validity in interpreting results without applying a required statistical test, and therefore, researchers should base their inferences on a statistical test or analysis when they publish an outcome. It is not enough just to select the correct test and to give the correct name in studies. As stated in Strasak et al. (15), when using more than one statistical test or technique, it is also necessary to specify which test is used for which data.

When we look at the rates of other kinds of errors, mathematical notation errors and statistical symbol errors were 13.38% and 3.18% for this study, respectively. These errors were reported as 6.93% and 3.23% in medical journals, and 2.94% and 3.43% in veterinary journals, respectively (12, 13). These results indicate that researchers who publish in radiology journals are lacking in knowledge of mathematical notation and statistical symbols, or that they do not take the necessary care in this regard. In 19.11% of the articles, there were incomprehensible statistical terms. Ercan et al. (12, 13)
found this kind of error in studies in medical journals as 4.15% and in veterinary journals as 0.49%, but it is remarkable that this error is more commonly encountered in radiology journals. Similarly, the rate for statistical terminology errors was 19.11% in radiology journals. Inappropriate interpretations were found in 8.92% of the examined articles. In medical journals the rate of this kind of error was reported as 8.76%; and in veterinary studies as 14.71% (12, 13). The rate of presentation of the statistical method, analysis and results in the incorrect section of the manuscript was 9.55% in our study. In order to achieve high quality in every aspect of a scientific study, particular attention should be paid to correct scientific notation, presentation, and expression as well.

In our study we also tested whether the statistical errors differed according to the rank of the journals in a well-known and commonly used ranking list. Taking the IF into consideration, there was no statistically significant difference between the groups with IF ≥2 and IF <2. With the increase of IF, the prestige of the journal increases, but it does not seem to have any effect on the decrease of statistical errors. This result shows that even though the IFs are high, efforts should be made to avoid statistical errors in publications in scientific journals and to increase the correct use of statistics.

According to the results of this study, statistical errors are frequently observed in radiology journals as well. Among the main reasons for these errors made by researchers are not consulting a biostatistics specialist about the subject, assuming that they know about statistics very well but in fact not having enough knowledge, and being careless (12, 32). Prevention of statistical errors in manuscripts is the responsibility of both the researchers who conduct scientific studies and the editors who publish these studies in their journals.

A researcher should have basic statistical knowledge to be able to read and interpret statistical methods in a scientific study. In order to ensure acquisition of statistical literacy, it must be taken into account that statistics should be taught accurately and adequately to medical students and to those receiving residency training (29). In addition, Altman has made suggestions in particular to develop standard education in statistics (33). Giving the necessary importance to statistics education will prevent the student from making serious mistakes in future scientific research, and it will ensure that the student will have enough statistical literacy. Another suggestion is to encourage the learning of these topics through seminars on critical thinking skills and research methodology in scientific meetings in areas other than medical training (29).

It is very important that the hypothesis of a study is designed so that it can be adequately assessed, that the data is collected appropriately and that the collected data is correctly analyzed. In this context, it will be useful to receive statistical consultancy at all stages of the study such as the planning, execution, data collection and analysis stages (9). Researchers should include biostatistics specialists in their scientific study and should take advice from them before they start a scientific study and at all following stages of the study (2, 33).

The greatest responsibility for journal editors is to be more sensitive to statistics in the article review process and to request the
help of biostatistics specialists as well as experts in the relevant topic in the process of evaluating scientific publications submitted to their journals. The reviewers who criticize the articles in scientific journals are usually selected according to their expertise in the relevant medical subject, and as a result, the statistical method used in most researches may not have sufficiently detailed examination and evaluation (33). Goodman et al. (34) proposed that a reviewer pool could be established for the evaluation of methodology in scientific studies and that journals could select referees from this pool. Altman (6) also discussed the effects and importance of having a statistics reviewer for scientific studies. It would be a good idea to include biostatistics specialists in this kind of pool for methodology in the light of these two remarks. Today, journals that use biostatistics reviewers and have biostatistics specialists on the editorial board are increasing in number (33). All scientific journals are expected to adopt this practice. It should not be forgotten that misuse of statistics may lead to misleading results, poor science, and in the end, to inappropriate patient care (35).

One of the subjects that may be useful in the review process is that of performing a statistical review before other relevant field experts' reviews (13). For, a mistake that is found in the result of a statistical review and needs to be corrected will affect the results and consequently the discussion of the study, and therefore, the statistical review should be carried out before the review by the relevant experts in order to avoid unnecessarily prolonging the article review process.

In addition to all these, the use of guidelines that have been agreed on by journal editors on the prevention of statistical errors in articles may be made more widespread. Some journals publish statistical guidelines in this regard, while others produce statistical checklists for referees. In addition, journals may include special statistical sections and series to draw attention and educate researchers in order to prevent statistical errors. Such methods may be useful to authors, in designing studies and analyzing their data; to reviewers, in the evaluation and criticizing of manuscripts; and to readers in understanding and interpretation of the published articles (25). These methods can contribute to the increase of statistical and scientific quality of publications.

Our study has some limitations. The articles we examined are in journals indexed in SCI or SCI-E in the Thomson Reuters Clarivate Analytics database. This study can be extended to include other radiology journals. However, the results of this study show that statistical errors are encountered even in well-known radiology journals. Topics such as study design and sampling were excluded from this study. Furthermore, statistical error classification does not take into account the severity of these errors and their potential consequences.

In conclusion, radiology journals, as do journals in different fields, include articles containing statistical errors. Statistical error rates are similar between the higher impact and lower impact radiology journals. Prevention of statistical errors in manuscripts is the responsibility of both researchers who conduct scientific studies and editors who publish these studies in their journals. Researchers should have a basic statistical knowledge, and the editor must submit all manuscripts for a statistical review.

Conflict of interest disclosure
The authors declared no conflicts of interest.

References