**Anna Justyna Milewska**
**Robert Milewski**
**Urszula Górska**
**Dorota Jankowska**
Department of Statistics and Medical Informatics,
Medical University of Bialystok

# STATISTICAL METHODS IN POLISH MEDICAL PUBLICATIONS

**Abstract**: Conducting research in the field of medicine today requires knowledge of statistical tools. For various reasons their correct selection is often a difficult task. This paper summarizes the most commonly used statistical methods in Polish medical journals published in 2009. We studied whether the choice of statistical tools and the methods of their implementation is connected with the number of points awarded for particular journals by MNiSW.

## Introduction

Currently there are almost 10 thousands titles on the bulleted list of journals of the Ministry of Science and Higher Education (MNiSW). Nearly 250 of them refer to medical aspects and are published in Poland. Only 16 of them are characterized by the Impact Factor indicator, that highlighting shall entail an evaluation of at least 10 points. Most journals are assigned 4 points (104), 2 points (67) and 6 points (45). The score is determined on the basis of the opinion of experts appointed for this purpose by the Minister of Science and Higher Education. Next to this classification there are many other lists evaluating the quality of journals. It is worth mentioning the ISI Master Journal List, where at the beginning of 2010 were 247 Polish titles, including 55 medical journals. One may encounter an incorrect belief that journals on this list are those that have Impact Factor, which is not the rule. However, it is a collection of the best journals in their field. It is observed that each year the number of Polish journals in this segment is growing [1]. An interesting list is also published on the Internet platform Index Copernicus. Journals which are included here are also subject to eva-

ISBN 978–83–7431–266–0 ISSN 0860-150X 

luation, which also involves granting of points. We found 685 Polish entries, 481 of them being medical journals.

Scientific investigators should take into consideration the fact that they will eventually want to publish their findings. Unfortunately, not many take this into account and only when all the results of the statistical analysis are confirmed they note that the study could have been carried out otherwise. These problems may be associated with too small a sample, incorrectly chosen and unrepresentative sample, ambiguously worded questions, etc. Even before conducting research, the investigator must take into account what tools can used to analyze the results. One should begin by setting a specific purpose for which the research is to be used. The examination should be planned in most detail and clarity [2, 3]. The researcher should remember that "there are only a handful of ways to do a study properly but a thousand ways to do it wrong" [4]. It is clear that practically we are not able to determine the target group we are interested in. Therefore, we should choose a sample which will be a subject of the research. This group should be representative and selected at random. Information about each of its elements should be collected in the form prepared by us on the computer. It is noteworthy that even at this early stage of the examination we should use the knowledge of statistics [5]. Mistakes in this phase are difficult to fix later. Inexperienced researchers repeatedly formulate unclear questions. Sometimes a respondent does not give an answer to a question because he/she does not understand the idea of the author. This may lead to the rejection of the questionnaire from the analysis. If this situation occurs repeatedly and the question is important for the researcher, the examination becomes useless. When you select features that will be analyzed you should choose in what scale they will be described. This affects the choice of statistical tests used later. After preparing research and collecting information about the observation we proceed to develop statistical material. We organize gathered information using the statistical series, then the data are prepared for statistical analysis. Two groups of methods must be distinguished: descriptive statistics and statistical inference. Now, taking the used scale, sample size and other factors, the investigator has to chose an adequate tool from those available. This choice is virtually never obvious. Although the computer performs calculations, the researcher has to indicate the appropriate statistical methods. Interpretation of results also belongs to the person conducting the survey. At present, during analysis, an appropriate computer program is essential. Optimistic is the fact that access to commercial statistical packages is spreading, leading to increased standards of statistical studies [6].

However, lack of statistics knowledge makes useless even the best computer programs.

Authors of scientific publications present the results of their experimental work with more or less complex statistical procedures. Unfortunately, an average doctor, for whom mathematics and statistics are not close, when reading ignores parts of the text which are not important according to his/her opinion. This results in a vicious circle. The person reading a scientific paper who is not paying attention to applied methods, writing his/her own paper does not select them with the due diligence. Often one meets scientific publications which at first sight use certain statistical procedures, however, when the reader wants to learn what methods were used in this particular situation it turns out that is often impossible. Authors often include information suggesting the use of a statistical test, writing for example "statistically significant differences were found at $p < 0,05$". However, it is not specified how it has been designated. The only information in the text referring to the use of statistical methods is the name of the computer program. It should also be noted that one can still find articles where conclusions are not supported by applying methods of statistical inference.

One may meet with the opinion that you can use statistics to prove any hypothesis which the author of a study puts. The fact is that if we do not use statistical tools properly, for example miss important assumptions, we can get surprising conclusions, but not always true. Impartially conducted statistical inference does not lead to such errors. To make correct analysis one should plan in advance the whole investigation. It is important that the group which is the subject of research has to be selected at random and its size can not be too small. Important is also the choice of tests which are used. If one reads a number of publications one can have a feeling that some authors have their favourite methods and apply them regardless of the situation. Particular attention should be paid here to the frequently used Student's t-test and ANOVA [7]. Textbooks describing these methods give an assumption that data may be modified in such a way that it is possible to correctly carry out the analysis. One of them is the normality of distribution of examined variables. If this condition is not met there is the necessity to use alternative statistical methods. Unfortunately, not all authors are announcing whether they considered assumptions of the used tests [8, 9]. You can also meet interesting situations, but absolutely incorrect, where the assumption of normal distribution is actually checked, but another important condition for abundance is missed. Compatibility with normal distribution is checked using an adequate test and the abundance of the sample is less than 10. With such a small number of observations checking this assumption is not appropriate. Applying Student's t-test in this situation should not take place.

Today scientific research and the publication of articles requires knowledge of at least basic statistics. It seems illogical to expend funds and spend time on research to interpret the results in an incorrect way. For many years, in numerous publications, attention has been paid to the problem of improper use of statistical methods. Already in 1994 Douglas Altman in the article "The scandal of poor medical research", wrote that very often available techniques are used wrongly. In result, this leads to many methodological errors. Despite the passage of time we can observe that this situation is still up-to-date.

## Purpose of study

The objective of this study was the presentation of statistical methods most frequently used to analyze medical data and to analysis errors in statistical inference. Assessment of the above issues had been made in the context of the points awarded for journals by MNiSW.

## Material and methods

For statistical analysis 10 Polish medical journals were chosen at random. As a result of the draw it turned out that they were in the following groups: 3 journals were rated by MNiSW for 2 points, 4 for 4 points, 2 6 points and 1 for 10 points. We analysed all research articles which appeared in these journals in 2009 (n = 248). Articles assessed for 2 points were n = 27, 4 points n = 104, 6 points n = 90 and 10 points n = 27. Breakdown of the percentage of articles in this study considering the MNiSW score presents Figure 1.

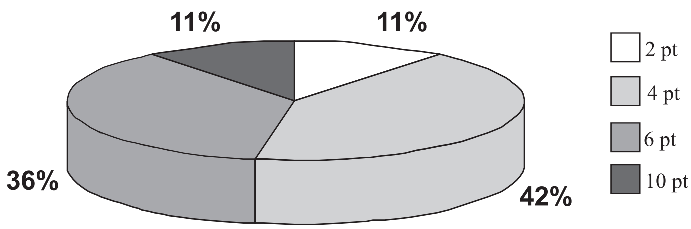


**Fig. 1. Breakdown of the percentage of articles based on the score of the journal**

In the present analysis for a small research or control group called for a group size of less than 30 items, and other of size 30 and more was large. Figure 2 presents the distribution of the sum of number of elements in the research and control group.

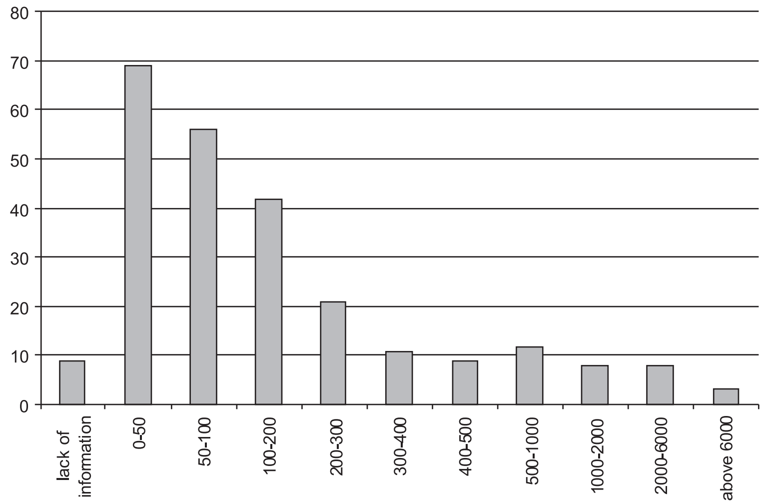


**Fig. 2. The number of elements including research and control group**

In statistical calculations to verify research hypotheses for qualitative variables the Chi-square test for independence was used. To the ordinal variables analysis and quantitative variables without normal distribution the following test were used: the Mann-Whitney U test, Kolmogorov-Smirnov test and Wald-Wolfowitz test for comparing two groups and non-parametric ANOVA Kruskal-Wallis test, median test and Post-Hoc multiple comparisons of average rang for comparing more groups.

The results were statistically significant at $p < 0,05$. For the statistical analysis the program Statistica 8.0 (StatSoft Inc.) was used.

## Results and discussion

During the analysis an alarming phenomenon of lack of statistical inference in the research articles was noticed (Table 1). No inference applies to 25% of all articles, the biggest problem exists in the journals for 4 points, there is no inference performed until 43% of the articles. By contrast, in journals for 10 points, always (100%) statistical inference was carried out. There is a significant statistical dependence between the fact of using me-

**Table 1**

**General characteristics of the analyzed research articles divided into MNiSW scores**

| Characteristic | 2 pt MNiSW n (%) | 4 pt MNiSW n (%) | 6 pt MNiSW n (%) | 10 pt MNiSW n (%) | Total n (%) | Statistical significance |
|---|---|---|---|---|---|---|
| Analyzed research Works | 27 (11%) | 104 (42%) | 90 (36%) | 27 (11%) | 248 (100%) | – |
| Works, in which applied statistical inference | 22 (81%) | 59 (57%) | 78 (87%) | 27 (100%) | 186 (75%) | $p < 0,001$ |
| Research group | | | | | | |
| – Small | 1 (4%) | 17 (16%) | 13 (14%) | 13 (48%) | 44 (18%) | $p < 0,001$ |
| – Large | 23 (85%) | 82 (79%) | 76 (84%) | 14 (52%) | 195 (79%) | |
| – Lack | 3 (11%) | 5 (5%) | 1 (1%) | 0 (0%) | 9 (4%) | |
| Control group | | | | | | |
| – Small | 1 (4%) | 4 (4%) | 12 (13%) | 0 (0%) | 17 (7%) | NS |
| – Large | 2 (7%) | 9 (9%) | 21 (23%) | 0 (0%) | 32 (13%) | |
| – Lack | 24 (89%) | 81 (88%) | 57 (63%) | 27 (100%) | 199 (80%) | |

thods of statistical inference in the articles of the journals and the number of MNiSW points that this journal has ($p < 0.001$). Median values in the group scoring MNiSW which used statistical inference is 6 points. And in the group in which these methods are not used is 4 points. Fromm [10] based on research conducted in the years 1982–1993 suggested that 13% of the articles was characterized only by using descriptive statistics (lack of statistical inference).

The analysis of the size of research groups indicates a significant relationship between group size and the number of points awarded by MNiSW for journals ($p < 0.001$). Small groups are used mainly for journals for 6 points (median), and large groups in journals at 4 points (median). Z. Sych in his work [11] observed a significant increase in the incidence of large samples (between 1988–1990). The inverse of this situation has been observed in the present study which explains better planning of medical research. The better the journal the better articles are published in it. This is often associated with more expensive research. A small study group appears in the 48% of articles for 10 points, and only 4% for the articles for 2 points. In turn, a large research group is up in 84–85% of the journals for 2 and 6 points.

The control group was found only in 20% of the analyzed articles. Most of the articles for 6 points – 36% of articles used in the control group. However, no control group was used in the articles for 10 points.

The presentation of the results using descriptive statistics occurred in almost all research papers (96%). The most common parameter used was the mean 41%–96%, standard deviation 29%–81%, range 7%–21% and a median 0%–20% (Table 2). There is a relationship between the use of mean ($p < 0.001$) and standard deviation ($p = 0.001$) and the points obtained by the journal from MNiSW. The median in the group using these parameters is 6 points, and the group is not applying the 4 points.

In 8% (n = 19) of the articles does not appear the name of the statistical test and in the results are given p-value. In this case, significant dependence on the MNiSW scoring is not observed, however, in articles for 10 points this problem does not appear.

Already in 1980, S. A. Glantz has observed in his study that the Student's t-test is the most popular and most widely used statistical test in medical articles [7]. In this study, it is used in 11%–48% of articles (Table 2) and significantly more frequently in journals bulleted higher ($p = 0.001$) – median 6 pkt. D. G. Altman in his work [9, 12] draws attention to the mistakes made in popular statistical analysis. In the case of Student's t-test the problem is usually that the data are not in line with the statistical assumption that both sets come from a population with normal distribution and have the same variance. In the present study, information about checking the normal distribution when applying the Student's t-test appeared only in 33% (n = 16) of analyzed articles. Checking the normal distribution should be done also in case of the analysis of variance and Pearson's correlation. Table 2 shows the frequency of checking of this assumption.

The analysis of variance (ANOVA) was used in 15% of the articles. There is a statistical relationship ($p < 0.001$) between the frequency of the used analysis of variance and MNiSW points. It has been used frequently in journals with higher scores (median 6 points).

Non-parametric tests were used in 43% of all articles, in particular groups their number ranged from 8 (which represents 30% of the 10 points group) to 45 times (representing 50% of the 6 points group). Although the application of nonparametric tests should be used in position location measurement (quartiles), median was given only in 18% (n = 19) of articles. However, most often the mean (61%) and standard deviation (54%) was presented, which should be presented when using parametric tests. L. Zaborski also notes that authors often do not examine the characteristics of normal distribution, or completely without knowing it, give to characterize the study population mean and standard deviation [8].

One of the most commonly used tests, in addition to Student's t-test, is the Chi-square test for independence [10]. In the present study it was used

**Table 2**

**The statistical methods used in analyzed research articles including MNiSW scoring**

| Statistical method | 2 pt MNiSW n (%) | 4 pt MNiSW n (%) | 6 pt MNiSW n (%) | 10 pt MNiSW n (%) | Total n (%) | Statistical significance |
|---|---|---|---|---|---|---|
| Descriptive statistics: | | | | | | |
| – Mean | 18 (67%) | 43 (41%) | 55 (61%) | 26 (96%) | 142 (57%) | $p < 0,001$ |
| – Standard deviation | 16 (59%) | 30 (29%) | 48 (53%) | 22 (81%) | 116 (47%) | $p = 0,001$ |
| – Median | 3 (11%) | 5 (5%) | 18 (20%) | 0 (0%) | 26 (10%) | NS |
| – Scope | 2 (7%) | 15 (14%) | 19 (21%) | 4 (15%) | 40 (16%) | NS |
| – Quartile | 0 (0%) | 1 (1%) | 6 (7%) | 0 (0%) | 7 (3%) | NS |
| – OR | 1 (4%) | 2 (2%) | 8 (9%) | 0 (0%) | 11 (4%) | NS |
| – SEM | 0 (0%) | 0 (0%) | 3 (3%) | 5 (19%) | 8 (3%) | $p < 0,001$ |
| No test name, and there is p-value | 2 (7%) | 6 (6%) | 11 (12%) | 0 (0%) | 19 (8%) | NS |
| Student's t-test | 5 (19%) | 11 (11%) | 20 (22%) | 13 (48%) | 49 (20%) | $p = 0,001$ |
| ANOVA | 2 (7%) | 9 (9%) | 13 (14%) | 12 (44%) | 36 (15%) | $p < 0,001$ |
| ANOVA + post hoc | 0 (0%) | 4 (4%) | 5 (6%) | 6 (22%) | 15 (6%) | $p = 0,004$ |
| Pearson correlation | 5 (19%) | 4 (4%) | 8 (9%) | 9 (33%) | 26 (10%) | NS |
| Check for normal distribution: | | | | | | |
| – Student's t-test | 1 (20%) | 1 (9%) | 11 (55%) | 3 (23%) | 16 (33%) | NS |
| – Analysis of variance | 2 (100%) | 0 (0%) | 5 (38%) | 2 (6%) | 9 (25%) | NS |
| – Pearson correlation | 3 (60%) | 0 (0%) | 1 (13%) | 2 (22%) | 6 (23%) | NS |
| Nonparametric tests | 16 (59%) | 38 (37%) | 45 (50%) | 8 (30%) | 107 (43%) | NS |
| Descriptive parameters used in the nonparametric tests | | | | | | |
| – Mean | 13 (81%) | 15 (39%) | 30 (67%) | 7 (88%) | 65 (61%) | NS |
| – Standard deviation | 11 (69%) | 12 (32%) | 28 (62%) | 7 (88%) | 58 (54%) | NS |
| – Median | 3 (19%) | 2 (5%) | 14 (31%) | 0 (0%) | 19 (18%) | NS |
| Mann-Whitney U test | 5 (19%) | 8 (8%) | 25 (28%) | 0 (0%) | 36 (15%) | NS |
| Chi-square test for independence | 7 (26%) | 26 (25%) | 20 (22%) | 2 (7%) | 55 (22%) | NS |
| – Chi-square with the amendment | 1 (14%) | 6 (23%) | 10 (50%) | 0 (0%) | 17 (7%) | NS |
| – Lack of necessary amendments | 5 (71%) | 14 (54%) | 6 (30%) | 1 (50%) | 22 (9%) | NS |
| Wilcoxon test | 2 (7%) | 1 (1%) | 6 (7%) | 3 (11%) | 12 (5%) | NS |
| Kruskal-Wallis test | 3 (11%) | 7 (7%) | 6 (7%) | 0 (0%) | 16 (6%) | NS |
| Spearman correlation | 1 (4%) | 7 (7%) | 8 (9%) | 3 (121%) | 19 (8%) | NS |

in 22% of the articles. Important is that this test with small abundance in subgroups has some modifications: Yates amendment, Fisher's exact test and the V-square test. In the current study, 9% of the articles did not apply "corrections" to the test despite the small abundance in subgroups [13].

Low levels of statistical methods used in the published articles observed in this study have long been a subject of particular concern. The basic condition for improving this situation is to improve the level of statistics and increase the accountability of journals.

D.G. Altman in his article "Improving the quality of statistics in medical journals" [9] put forward concrete proposals to raise the level of publications. First of all, he believes that all the research using statistical methods should be reviewed by the statistics. Subsequently, the revised articles should be returned to the same reviewer to re-evaluate. An interesting solution is the suggestion that the journal should supply statistical guidelines for the authors which should be the standard of all research and should also contain a separate paragraph devoted to statistical methods implemented.

## Conclusions

Journals with higher MNiSW scoring have a more valid statistical methods and more often use parametric methods such as the analysis of variance and Student's t-test. Large research groups are more often used in the articles for a smaller number of points MNiSW. Greater emphasis should be put on teaching biostatistics in medical studies and related fields to achieve improvements in the application of statistical methods. Journal editorial teams should employ statistical reviewers with experience.

## R E F E R E N C E S


[1] Racki G. Dwuznaczny urok listy czasopism punktowanych. W: Bibliograficzne bazy danych: kierunki rozwoju i możliwości współpracy. Ogólnopolska konferencja naukowa z okazji 10-lecia bazy danych BazTech. Bydgoszcz, 27–29 May 2009.

[2] Krzych J. Ł. Interpretacja wyników analizy statystycznej danych. Kardiochirurgia i Torakochirurgia Polska, 2007; 4 (3): 315–321.

[3] Moczko J. A., Brębowicz J. A., Tadeusiewicz R. Statystyka w badaniach medycznych. Wydawnictwo Springer Warszawa 1998.

[4] Sacket D. L. Rational therapy in the neurosciences: the role of the randomized trial. Stroke 1986; 17: 1323–1329.

[5] Gellerstedt M. Istotne znaczenie statystyki w badaniach medycznych. Alergia Astma Immunologia, 2003; 8 (1): 25–32.

[6] Zejda J. E. Medyczny artykuł naukowy. Zasady dobrej praktyki publikacji. Ann. Acad. Med. Siles. 2006; 60,4: 323–329.

[7] Glantz S. A. Biostatistics: how to detect, correct and prevent errors in the medical literature. Circulation, 1980; 61, 1.

[8] Zaborski L. Zastosowanie metod statystycznych w badaniach i błędy we wnioskowaniu (kłamstwa statystyczne). Medycyna Pracy XLVI, 1995; 3, sup. 4: 81–92.

[9] Altman D. G. Statystyka i etyka w badaniach naukowych. W: Statystyka w praktyce lekarskiej. Wydawnictwo PWN, Warszawa 1997.

[10] Fromm B. S., Synder V. L.: Research design and statistical procedure used in the Journal of Family Practice. J. Farm. Pract. 1986; 23: 564.

[11] Sych Z. Ocena stosowania metod statystycznych w wybranych krajowych czasopismach medycznych. Roczniki PAM w Szczecinie, 1996; 42: 67–84.

[12] Altman D. G. The scandal of poor medical research. BMJ 1994; 308: 283–284.

[13] Stanisz A. Przystępny kurs statystyki z zastosowaniem STASISTICA PL na przykładach z medycyny. Tom 1. Statystyki podstawowe. Wydawnictwo StatSoft. Kraków 2006.