Państwowa Medyczna Wyższa Szkoła Zawodowa w Opolu

ISSN 2080-2021



# THE USE OF OBJECTIVE STATISTICAL PROCESSING FOR PRESENTING THE DATA IN SCIENTIFIC PUBLICATIONS

## Stosowanie obiektywnej analizy statystycznej w celu prezentacji danych w publikacjach naukowych

OLGA FEDORTSIV[B,E,F]
NATALIIA LUCHYSHYN[B,E,F]

I.Ya. Horbachewsky Ternopil State Medical University of Ministry of Health of Ukraine, Ternopil

**A –** przygotowanie projektu badania | study design, **B –** zbieranie danych | data collection, **C –** analiza statystyczna | statistical analysis, **D –** interpretacja danych | data interpretation, **E –** przygotowanie maszynopisu | manuscript preparation, **F –** opracowanie piśmiennictwa | literature search, **G –** pozyskanie funduszy | funds collection

**SUMMARY**

The most important components of the quality of scientific publications are correct applications of statistical methods for the analysis of the phenomena and its completeness. Underestimation of this fact causes incomplete or incorrect descriptions and proves the research as superficial or insufficiently reliable. The minimum of necessary statistical information should be present in the publication. It allows the reader to make some comparisons. The use of certain rules of presentation of scientific and clinical information, regular open discussion and expertise of scientific publications serve as tools of self-initiation in the methodology of biomedical research. Visual elements help authors present detailed results and complex relationships, patterns and trends clearly and concisely, reducing the length of the manuscript and enhancing readers' understanding of the study results. The purpose of this paper is to highlight practical recommendations and to summarize relevant points of study analysis and reporting by using objective statistical processing for presenting data in scientific publications.

**Keywords:** statistical information, graphical data presentation, publication

**STRESZCZENIE**

Najważniejszymi elementami jakości publikacji naukowych są: poprawne stosowanie metod statystycznych do analizy zjawisk oraz ich kompletność. Nieuwzględnienie tego istotnego faktu skutkuje niepełnymi i niepoprawnymi opisami oraz sprawia, że badanie jest powierzchowne i niewystarczająco wiarygodne. Publikacja powinna zawierać niezbędne minimum informacji statystycznych, tak aby pozwolić czytelnikowi na dokonanie pewnych porównań. Zastosowanie określonych zasad w prezentowaniu danych naukowych i klinicznych pozwala wykorzystać artykuły naukowe w otwartych dyskusjach i na szkoleniach. Wizualne elementy mogą pomóc autorom przedstawić jasno i zwięźle szczegółowe wyniki oraz złożone relacje, wzorce i trendy, skracając rękopis oraz powodując lepsze zrozumienie wyników badań u czytelników.

Celem pracy jest przedstawienie konkretnych zaleceń dotyczących stosowania obiektywnego statystycznego przetwarzania dla prezentacji danych w publikacjach naukowych na podstawie doświadczeń autorów i recenzentów, jak również redaktorów czasopism.

**Słowa kluczowe:** informacja statystyczna, graficzne przedstawianie danych, publikacja

## Background

The final result of the expensive medical research is a couple of publication's pages in a professional medical journal. The most important components of the quality of scientific publications are correct application of statistical methods for the analysis of the phenomena and its completeness. Underestimation of this fact causes incomplete or incorrect descriptions and shows research as superficial or insufficiently reliable. Scientists engaged in biomedical research mostly use statistical language such as English. Scientific publications grant to authors of the article priority of the results, but also allow the reader to compare the findings of the publication with their own results. That is why minimum of necessary statistical information should be present in the publication. It allows the reader to make such comparisons. In addition, articles have intrinsic value as the material for training for many readers. It requires a clear and complete account of the methods of analysis and interpretation of research results [1].

As long as there are no developed training programs for physicians in clinical Epidemiology and Biostatistics, they can only rely on the self-study in these fields. It is well known that without motivation it is impossible to provide quality in the self-learning process. The use of certain rules of presentation of scientific and clinical information, regular open discussion and expertise of scientific publications are the tools of self-initiation in the methodology of biomedical research [2,3]. Visual elements help authors present detailed results and complex relationships, patterns and trends clearly and concisely, reducing the length of the manuscript, and enhancing readers' understanding of the study results [4].

The purpose of this paper is to highlight practical recommendations for the use of objective statistical processing for presenting the data in scientific publications based on our and the reviewers' experience as well as journal editors. Below we have summarized relevant points regarding study analysis and reporting.

The guiding principle for statistical reporting is recommended by The International Committee of Medical Journal Editors. The main statement of Uniform Requirements for Manuscripts submitted to biomedical journals notice that each publication reporting statistical analyses should describe statistical methods with enough detail and provide enough detail of results for incorporation into other analyses [5].

According to the recommendations an author should specify the name of a statistical test, total sample size (n) in each of the statistical analysis and the rationale for selection of a particular test. Actual values «p» should be cited in each test (not just «significantly» or «p <0.05»). It should be clearly stated which statistical test was used for this p-value. In case of short manuscript such information should be reported in the text, or be reflected in the figures and captions. Sample data should be presented in the form of descriptive statistics, which should include: the sample size (n) for each data set; measures of spread such as standard deviation and span. Span is more acceptable for small samples size than the standard deviation. The graphs should clearly display the limits of error or confidence intervals. Authors should indicate whether the number following the symbol $\pm$, is an standard error of the mean (SEM) or standard deviation (SD). Authors are obliged to justify the application of specific criteria. They ought to define statistical terms, abbreviations, and most symbols as well as specify the computer software used [6,7].

Qualitative data (nominal and ordinal) is also called categorical data. This data is generalized by counting the number of observations in each category (frequency) or calculating the proportion of observations in each particular category of the studied sample (relative frequency, percent frequency). If the author uses relative values she/he should also specify the absolute frequency of a certain value by the total number of data. Percentages are useful to compare groups with different number of variables. The raw data should be shown this way especially when used for calculation; for example: 53.3% (16/30), 33% (30 out of 90 patients). In case of iterative percentage recalculation in the subgroup in relation to the group (percent of a percent), this procedure should be consistently described in detail. If the sample size is greater than 100, it is recommended to specify the percent with not more than one decimal place. If the sample size is less than 100, the percent should be indicated by only integers. If the sample size is less than 20, the percent is generally not recommended. In such cases only the absolute number of the total sample should be used.

Quantitative variables are variables measured on a numeric scale. Height, weight, response time, subjective rating of pain, temperature, and score on an exam are all examples of quantitative variables. Description of quantitative data depends on the type of distribution. For continuous data that is close to a normal distribution, means and standard deviations (SD) (not standard errors) should be applied. For non-normal, skewed data such as duration, BMI etc. author should give medians and upper and lower boundaries of interpercentile ranges and the minimum and maximum values of ranges, not just the size of the range. It is not recommended to use the standard error of the mean (SE) to indicate the variability of a data set [8]. Absolute numbers and percentages must be given to count data, in particular for small studies. When sample size is below about 200, percentages should be given without decimal places. Otherwise one decimal place is usually sufficient. Report only the relevant correlations, accompanied by a confidence interval.

Results of statistical analysis can be presented in the text, in a table, or pictorially as a chart, diagram or graph. But it is easier to understand and interpret data when they are presented graphically than using words. Tables and figures are an integral part of a well-written scientific paper, not an adjunct. The bulk of the detailed information in a paper is typically presented in its tables. Many of the descriptions and basic concepts, key discoveries, and some of the conclusions are presented in figures. The author should consider whether a figure or a table is more appropriate in article [9–11].

There are few recommendations, which are collected from big amount of reference and are cumulated in tips on effective use of tables and figures in research papers [12]. It recommends for authors to be sure that display items are self-explanatory, because some read-

ers pay their attention to the tables and figures before they read the entire text. The text should always mention the key points in a table or figure: if it does not warrant the discussion, it should not be there. The author should write the verbal summary before preparing the final version of the tables and figures and should make sure they illuminate important points. But it is important to refer to, but not repeat in the text, the details which are the key points of the table or figure. The values or details in a table and those in the text should be the same. When referring to tables and graphs from within the text, the author can use: clauses beginning with "as": "As shown in Table 1, ..."; passive voice: "results are shown in Table 1"; active voice: "Table 1 shows that ..."; parentheses: "Physical examination of children has revealed different levels of development (Table 1)".

The titles should be informative and concisely describe the purpose or contents of the table or figure. The column heads, axis labels and figure labels, etc. must be clearly and appropriately labeled. The author should follow journal guidelines and adhere instructions about numbering of tables and figures, the style of numbering, titles, image resolution, file formats, etc. [13–15].

Making decisions about how to present data is an important part of planning scientific papers, because some journals limit the number of tables and figures. Tables are better than graphs for giving structured numeric information, while the graphs are better for indicating trends and making broad comparisons or showing relationships [16].

It is recommended to use tables for presentation of many numerical data in a small space, for comparing of data values among related items and for detection of specific characteristics of data. Particularly useful suggestions are the combination of repetitive tables, dividing the data into clear and appropriate categories of tables, making the long tables a part of the appendix or supplemental material; use sufficient spacing between columns and rows [17].

## Examples of tables

**Table 1.** Frequency distribution of physical development level in children

| Level of Development | Total | | Male | | Female | |
|---|---|---|---|---|---|---|
| | n | % | n | % | n | % |
| Low | 26 | 15.3 | 14 | 14.4 | 12 | 16.4 |
| Average | 110 | 64.7 | 65 | 67.0 | 45 | 61.7 |
| High | 34 | 20.0 | 18 | 18.6 | 16 | 21.9 |
| Total | 170 | 100.0 | 97 | 100.0 | 73 | 100.0 |

Tables are easily constructed using word processor's table function or a spread sheet program such as Excel. The elements of a table include the Title, Column Titles, and the Table Body (quantitative or qualitative data). They may also include subheadings and footnotes. A well-organized table allows readers to grasp the meaning of the data presented with ease, while a disorganized one will leave the reader confused about the data itself, or the significance of the data.

In the paper for publication, tables should be labeled with a number preceding the table title, centered on the page, numbered in the order they appear in the text, referenced in the order they appear in the text. Do not use too many small tables for information that could be presented better in the text. It is also recommended to label the table number and descriptive title above the table, column and row labels that describe the data, including units of measurement. Tables should be set apart from the text itself (text does not flow around the table).

Instead of presenting large tables, presentation in graphic form is preferred to make it easier to grasp the key message. Using figures is recommended if the aim of the presentation is to summarize research results, to show explanation of a sequence of events, procedures, physical characteristic, to submit trends, relationships between data set, especially if general pattern is more important than the exact data values [18,19].

Figures are visual presentations of results, including graphs, diagrams, photos, drawings, schematics, maps, etc.

There are many types of graphs that can be used to portray distributions of quantitative variables. The choice of graph type (e.g., line, pie chart, bar, stacked bar, scatter plot,) depends on the author' experience, knowledge, and expectations.

Some graph types such as stem and leaf displays are best-suited for small to moderate amounts of data, whereas others such as histograms are best-suited for large amounts of data. Graph types such as box plots are good at depicting differences between distributions. Scatter plots are used to show the relationship between two variables [20].

Line graph may be used for depiction of relative or absolute amounts, the rate of increase in the means of the dependent variable as a function of changes in the independent variable (Figure 1).

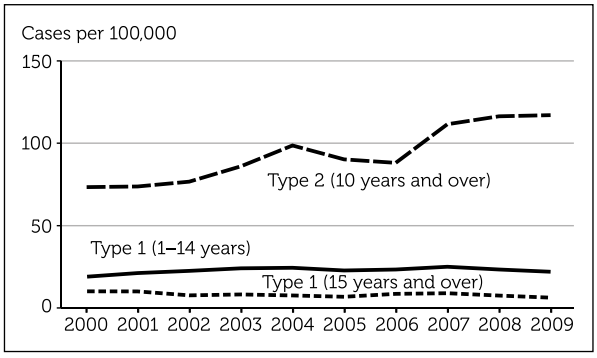


**Figure 1.** New cases of insulin-treated diabetes per 100,000 population, 2000 to 2009 (Source: 2009 National Diabetes Register (data extracted 2012).

Pie charts are effective for displaying the relative frequencies of a small number of categories. They are not recommended, however, when a large number of categories must be presented. Pie charts can also be confusing when they are used to compare the outcomes of two different surveys or experiments (Figure 2).

Bar graph is useful for determination of the difference between the means of the dependent variable across different levels of the independent variable. If the author needs to determine proportions but not absolute amounts, it is good to use a pie chart or divided bar graph (i.e., stacked bar graph). Scatter plot usage

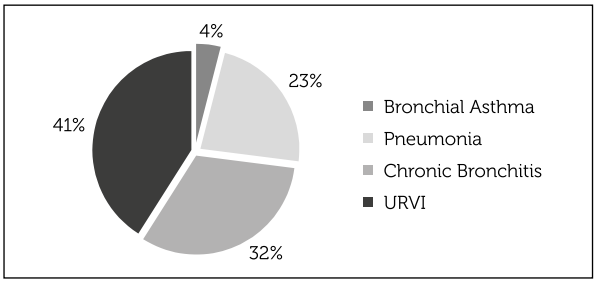


**Figure 2.** Pie chart of diseases illustrating frequencies it's prevalence among children.

is recommended if the researcher needs to determine the degree of correlation between two variables. Typically, the y-axis shows the number of observations in each category rather than the percentage of observations as typical in pie charts.

Several recommendations in figures creation such as all the parts of the figure should be clear with standard font; legible labels against the figure background; legends of the graphs and label the key sections (parts of diagrams, photographs, and all axes, curves, and data sets in graphs and data plots), should be used for the explanation of the key message. It is recommended to include scale bars in images and maps; to specify units wherever quantities are listed; appropriate use of relation of x- to y-axis, sufficiently thick lines, use of color instead of dot, dash-dot etc., and the axis should not be exaggerated to artificially inflate a minor difference, e.g. by truncating the axis (Figure 3) [21].

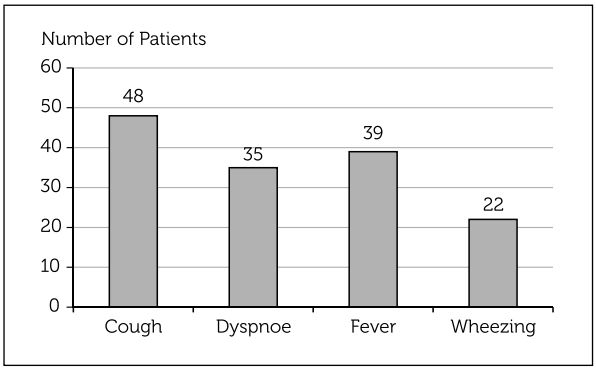


**Figure 3.** Bar chart of prevalence rates of respiratory symptoms in children

If the author needs to compare the results of different surveys, or of different conditions within the same overall survey, she/he could compare the "distributions" of responses between the surveys or conditions. Bar charts are often excellent for illustrating differences between two distributions [5]. Figure 4 shows the number of children having allergic disorders dependant on sex.

The author should pay attention to using similar patterns or shapes to indicate data from similar conditions in a study, and use different patterns or shapes for the indicators of conditions intended to be distinguished.

Stacked bar graph is a graph that is used to compare the parts to the whole. The bars in a stacked bar graph are divided into categories. Each bar represents a total (Figure 5).

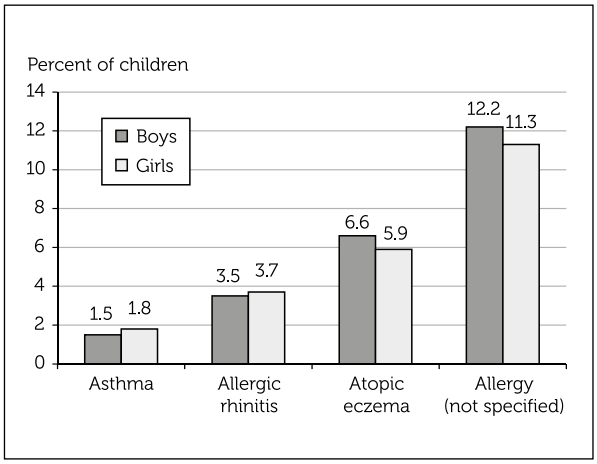


**Figure 4.** Sex distribution of allergic disorders

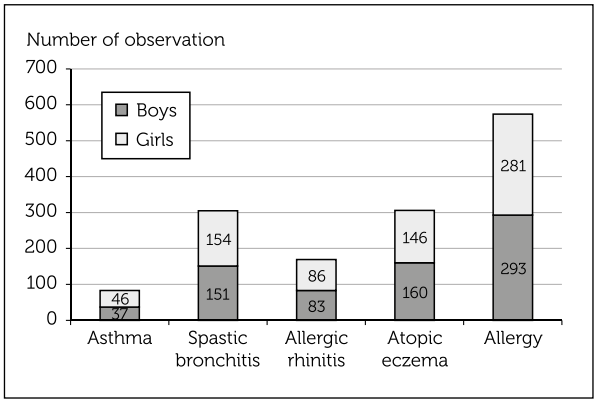


**Figure 5.** Sex distribution of allergic disorders

## Conclusions

Graphing data is the first and often most important step in data analysis. Many options are available for presenting data, especially with the increasing capabilities of wordprocessor, spreadsheet, and graphics software for creating tables and graphs. Well-presented tables and graphs can concisely summarize information which would be difficult to describe in the text itself.

The tables present lists of numbers or text and can be used to synthesize existing literature, to explain variables, or to present the wording of survey questions, raw data, not when you want to show a relationship between variables. They are also used to make a paper or article more readable by removing numeric or listed data from the text.

Figures are visual presentations of results, including graphs, diagrams, photos, drawings, schematics, maps, etc. The figures provide visual impact and can effectively communicate your primary finding. Traditionally, they are used to display trends and patterns of relationship, but they can also be used to communicate processes or display complicated data in a simple way. Figures should not duplicate the same information found in tables and vice versa.

## References

1. Baschinskiy SE. Some questions of journalistic ethics. *Cardiology* 1995; 6: 89–92.
2. Chan KBY, Man-Son-Hing M, Molnar FJ, Laupacic A. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. *Can Med Assoc J 2001*; 165(9): 35–39.
3. Vorobyov KP. Format of the modern journal publication of the clinical research. Part 4. Biostatistics. *Morphology* 2009; 3(4): 102–114.
4. Rodrigues V. Tips on effective use of tables and figures in research papers. [online] 2014 Sep [cited 18.10.2014]. Available from URL: http://www.editage.com/insights/tips-on-effective-use-of-tables-and-figures-in-research-papers#.
5. Orekhov LI, Karavayeva EL. The need for appropriate statistical and experimental methods to modern requirements. *Theory Pract Phys Culture J* 2005; 3: 46–49.
6. How robus are you data? New rules for the presentation of statistics. *Nature Cell Biology* 2009; 11: 667. doi:10.1038/ncb0609-667a
7. Guide to authors. Nat Med 2008; 10: 1–8 [online] 2014 Oct [cited 29.10.2014]. Available from URL: http://www.nature.com/nm/pdf/gta.pdf.
8. Lang TA, Secic M. *How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers*. 2rd ed. Philadelphia: ACP; 1997: 367.
9. World Association of Medical Editors. Recommendations of the World Association of Medical Editors on the principles of ethics of scientific publications in medical journals. Intern J Med Pract 2005; 5: 24–30 [online] 2013 Jul [cited 10.09.2014]. Available from URL: http://www.mediasphera.ru/journals/practik/169/2382/.
10. Lang TA, Altman DG . Basic Statistical Reporting for Articles Published in Biomedical Journals: the statistical analyses and methods in the published literature or the SAMPL Guidelines [online] 2014 Oct [cited 21.11.2014]. Available from URL: http://www.equator-network.org/wp-content/uploads/2013/07/SAMPL-Guidelines-6-27-13.pdf.
11. Chan KBY, Man-Son-Hing M, Molnar FJ, Laupacic A. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. *Can Med Assoc J* 2001; 165(9): 35–39.
12. Lang T, Altman D. Statistical analyses and methods in the published literature: the SAMPL Guidelines. [online] 2014 Oct [cited 30.11.2014]. Available from URL: http://www.equator-network.org/wp-content/uploads/2013/07/SAMPL-Guidelines-6-27-13.pdf
13. International committee of medical journal editors. Uniform requirements for manuscripts submitted to biomedical journals. *Ann Intern Med* 1997; 126: 36–47.
14. Greenacre MJ. *Theory and Applications of Correspondence Analysis*. London: Academic Press; 1987.
15. Online Statistics Education: an interactive multimedia course of study. [online] 2014 Oct [cited 28.11.2014]. Available from URL: http://onlinestatbook.com/2/index.html.
16. Milewska AJ, Jankowska D, Górska U, Milewski R, Wołczyński S. Graphical representation of the relationships between qualitative variables concerning the process of hospitalization in the gynaecological ward using correspondence analysis. *Studies Logic Grammar Rhetoric* 2012; 29 (42): 7–25.
17. Petrie A, Sabin C. Medical statistics at a Glance. 3rd ed. New Jersey: Willey Blackwell; 2009.
18. Sonnad SS. Describing data: statistical and graphical methods. *Radiology* 2002; 225(3): 622–628.
19. Simundic AM. Practical recommendations for statistical analysis and data presentation in Biochemia Medica journal. Biochemia Medica 2012; 22(1): 15–23 [online] 2014 Sep [cited 26.10.2014] Available from URL: http://dx.doi.org/10.11613/BM.2012.003.
20. Monsen ER, Van Horn L. *Research: Successful Approaches*. 3rd Ed. Chicago: American Dietetic Association; 2007.
21. Statistical guidelines. European Heart J [online] 2014 Nov [cited 9.01.2015] Available from URL: http://eurheartj.oxfordjournals.org/for_authors/stat_guide.html.

**Correspondence address:**
Nataliia Luchyshyn
Protasevycha St. 18/116
46-001 Ternopil, Ukraine
Tel. +380 066 047 11 20
E-mail: luna.ternopil@gmail.com

Received: 29.03.2015
Reviewed: 30.03.2015
Accepted: 31.03.2015