Invited Review / Davetli Derleme
doi: 10.18678/dtfd.1094422

Duzce Med J, 2022;24(1):1-6
Düzce Tıp Fak Derg, 2022;24(1):1-6

Conscious and Correct Use of Biostatistical Methods in Medical Researches: From Planning to Reporting the Results - Part I

Tıbbi Araştırmalarda Biyoistatistiksel Yöntemlerin Bilinçli ve Doğru Kullanımı: Planlamadan Sonuçların Raporlanmasına - Bölüm I

Erdem KARABULUT
0000-0002-7811-8215

Department of Biostatistics, Hacettepe University School of Medicine, Ankara, Turkey

Chair of the Biostatistics Association

ABSTRACT

The principles and methods of biostatistics serve as a guide for healthcare practitioners in both their daily and scientific work. Biostatistics principles and methods should be considered at every stage. A scientific research requires a multidisciplinary teamwork and each team must include a biostatistics expert at the planning stage of the study. During the planning phase, a biostatistics expert can answer all questions about the design, conduct, data processing, data analysis, presentation of the results, and publication of the results. Reporting guidelines for almost all research designs were developed to improve the quality and transparency of research reports. When the research results are presented in a structured and standardized way using these guidelines, it will allow the readers to interpret the findings more easily and accurately. It is now easier for researchers to gain access to a variety of software or web applications that allow them to perform basic and advanced statistical analyses. However, erroneous/misleading results are produced when these software are used unconsciously or incorrectly by those with insufficient statistical knowledge. Health professionals should also have sufficient statistical knowledge to identify errors in articles related to their field. Therefore, they are expected to have knowledge of research designs and basic statistical concepts. Knowing the basic concepts such as population, sample, sampling methods, random assignment of volunteers to groups, parameters, statistics, data types, etc. will help the readers while evaluating the material and method sections of the articles.
Keywords: Biostatistics; medicine; research; planning; reporting guidelines.

ÖZ

Biyoistatistik ilke ve yöntemleri, sağlık çalışanlarına hem günlük hem de bilimsel çalışmalarında rehberlik eder. Biyoistatistik ilke ve yöntemleri araştırmaların her aşamasında dikkate alınmalıdır. Bilimsel bir araştırma, multidisipliner bir ekip çalışmasını gerektirir. Her ekipte mutlaka bir biyoistatistik uzmanı bulunmalıdır ve çalışmanın planlama aşamasından itibaren ekibe dahil edilmelidir. Planlama aşamasında; araştırmanın tasarımı, yürütülmesi, verilerin işlenmesi, verilerin analizi, sonuçların sunumu ve yorumlanarak yayın haline getirilmesi aşamalarıyla ilgili tüm sorular biyoistatistik uzmanın yardımı ile cevaplanabilir. Araştırma raporlarının kalitesini ve şeffaflığını artırmak için hemen hemen tüm araştırma tasarımları için raporlama kılavuzları geliştirilmiştir. Bu kılavuzlar kullanılarak araştırma sonuçları yapılandırılmış ve standart bir şekilde sunulduğunda, bulguların okuyucular tarafından daha kolay ve doğru yorumlayabilmelerine olanak tanıyacaktır. Günümüzde araştırmacıların temel ve ileri düzey istatistiksel analizleri yapabilecekleri birçok yazılıma ya da web uygulamalarına ulaşımları kolaylaşmıştır. Ancak, istatistik bilgi düzeyi yetersiz olanlar tarafından bu yazılımların bilinçsiz ve yanlış kullanılmaları hatalı/yanıltıcı sonuçların elde edilmesine neden olmaktadır. Sağlık çalışanlarının kendi alanları ile ilgili makaleleri okurken yapılmış belirgin hataları belirleyebilecek düzeyde istatistik bilgisine sahip olmaları da gerekir. Bu nedenle, araştırma tasarımları ve temel istatistiksel kavramlar hakkında bilgi sahibi olmaları beklenir. Popülasyon, örneklem, örnekleme yöntemleri, gönüllüleri gruplara rastgele atama yöntemleri, parametre, istatistik, veri tipleri vb. temel kavramları bilmeleri, makalelerin materyal ve metot bölümlerini değerlendirirken okuyuculara kolaylık sağlayacaktır.
Anahtar kelimeler: Biyoistatistik; tıp; araştırma; planlama; raporlama kılavuzları.

Corresponding Author
Sorumlu Yazar
Erdem KARABULUT
ekarabul@hacettepe.edu.tr

Received / Geliş Tarihi : 03.02.2022
Accepted / Kabul Tarihi : 27.03.2022
Available Online / Çevrimiçi Yayın Tarihi : 29.03.2022

INTRODUCTION

The principles and methods of biostatistics serve as a guide for healthcare professionals in both their daily and scientific work. Conscious use of this guide will take health professionals in the right way, and unconscious use will lead to the wrong way.

There is a misconception among researchers that the biostatisticians were solely responsible for analyzing the data from the completed study. However, the task of biostatistics in research is to ensure that every study is organized, conducted, and completed according to scientific methods. Therefore, biostatistics principles should be considered at every stage from the planning of the study to the report writing.

It is against the nature of science for a person to be an expert in every subject, no matter how broad his knowledge and experience. Therefore, a scientific research cannot be a one-man operation. It requires a multidisciplinary teamwork. Each research should be planned, conducted, analyzed, and finalized by writing a report by a team of experts from different branches related to the study. Every research team must have a biostatistician and he/she will be included in the team at the beginning of the study. If this is not possible, biostatistics consultancy should be obtained from the beginning of the study.

Researchers perceive research and statistics as separate sciences and consider themselves as a researcher and a biostatistician as a person who will apply appropriate methods to the data they have collected. For this reason, he organizes the stages of the research (planning and data collection) that he considers to be of interest to him and compiles his data. Since he thinks that biostatistics methods will be used in the future, he applies to a biostatistics expert to evaluate the data he has collected or to apply some significance tests (1).

If the research is not well organized, in other words, it has no scientific value, applying even the most advanced biostatistics methods to this research will not bring a solution. Andrew Lang's (1905) quote explains this very well: “He uses statistics as a drunken man uses lamp-posts... for support rather than illumination.” (2). The contribution of a biostatistician to a research at the planning stage is more important than what he/she will do at the evaluation stage.

The unconscious and wrong use of statistics in published articles is a problem that we constantly encounter. Statistical errors can be made at any, some, or all stages of the planning of the research, designing, data collection, analysis of the collected data, the presentation of the obtained results with appropriate tables and graphics, and the interpretation of the results based on the research findings.

The most important stage of the research is the planning stage. At this stage, very important mistakes are encountered. The main reason for these mistakes is that a biostatistician was not included in the team at the beginning of the study or professional statistical consultancy was not received. The best time a statistician is needed is during the planning phase of the study. At this stage, it is much easier to identify and correct any deficiencies.

At the planning stage of the study;

- Forming a research question
- Creating research hypotheses in accordance with the objectives of the study
- Determining the research design (observational, experimental, meta-analysis, etc.)
- Determining what measure is most appropriate for measuring the phenomenon of interest, when and with what measuring instrument, who will measure it, etc.
- Defining the population and sample of the study
- Evaluation of the availability of a source from which data can be obtained
- Which statistical methods will be used to test the research hypotheses
- Calculation of the minimum number of subjects required for the study
- Whether the control group will be used in the study
- Which sampling method will be used in the study
- How to assign participants to groups
- Whether blinding will be used in the study, and if so, who will be blinded
- How to handle missing data problem
- How to handle outliers/extreme values
- Are there any ethical issues related to conducting the study
- What kind of tables and graphics will be utilized to summarize the findings (draft tables and graphs can be prepared)
- What interpretations and inferences will be made
- Who will be the author in the study, their responsibilities, and the order of the authors

etc. many more questions need to be addressed. All of these questions can only be answered with the help of a biostatistician. Statistics can be considered as the common language of all fields of science. It is also a fact that the results obtained from research processes in which this common language is not used appropriately are not accepted in the scientific community (3-5).

The following statement, cited by Altman (6) from Mainland (1950), highlights the importance of statistical principles in scientific research: “Finally, it must be stressed again that, whatever sources of help are found and whatever techniques are employed, the investigator himself has to grasp the principles of statistical reasoning... modern statistical principles are not something that we can take or leave as we wish, for they comprise the logic of the investigation in all fields, including the field of clinical research”.

Biostatistics education for health professionals at the undergraduate and graduate levels is critical. Because, the perspective of a health professional who understands these principles and methods adequately and can use them with skill expands on the individual and community events. As a result, health professionals develop the ability and skill of perceiving and thinking in multidimensional, rather than one-dimensional, ways. Health professionals are aware that the only way for the findings to be valid and reliable is the scientific approach, and the only factor that will ensure this is the appropriate and correct use of biostatistics principles and methods. This awareness gives health professional the ability and skill to always approach events with a scientific view, and to solve problems.

Today, the increase in the number of statistical package programs and the easy accessibility of these softwares offer researchers some advantages, but on the other hand, it has become an important problem. Researchers can utilize both univariate and multivariate statistical methods with applications available as free softwares or web pages. Statistical package programs have been developed to make mathematical calculations easier. The unconscious use of these packages by people with insufficient biostatistical knowledge leads to erroneous results. Hofacker (7) expressed this situation in his article as follows: "The good news is that statistical analysis is becoming easier and cheaper. The bad news is that statistical analysis is becoming easier and cheaper".

Various measures have been taken to reduce the sources of bias and statistical error in research as much as possible and to standardize the processes. In particular, Bradford Hill`s study titled "The time to allow for statistical factors is when an inquiry is planned, not when it is completed" published in The Lancet since the 1930s drew attention to this issue and statistical errors in studies published in scientific journals began to be examined. Although significant progress has been made over the years, it is observed that serious statistical errors continue to be made in scientific studies in recent years. In fact, the use of complex statistical methods is increasing, and this causes readers to be more disadvantaged (6,8).

REPORTING GUIDELINES

Editors/referees would benefit from a list of features to consider when evaluating a publication. However, depending on the type of research design, there are many different points to consider, and these can be overlooked. Furthermore, people's perspectives on various issues may differ depending on their areas of expertise. As a result, working groups comprised of journal editors, reviewers, authors, and other stakeholders developed reporting guidelines for all research designs in order to improve the quality and transparency of research reports (9).

These guidelines state point by point what to consider in the study title, abstract, introduction, material-method, results, and discussion sections. However, there are also problems with the use of these guides. Caulley et al. (10) examined 200 articles that reported using the guidelines for four different study designs and found that only 39% of those articles used the guidelines appropriately.

Almost 500 improved guidelines and their extensions can be accessed on the Enhancing the QUAlity and Transparency Of health Research (EQUATOR)-network (https://www.equator-network.org/) web page. A brief list of available guidelines for the main study designs is given in Table 1 (11).

It would be helpful to list the features to be considered while reading or reviewing a scientific publication. Ideally, an article should be expected to answer "yes" to all the questions on the list. But very few articles can fulfill this expectation. Naturally, not all questions on the list are of equal weight. The first consideration should be given to the design of the study and the risk of bias. Because, if the design of the study is unacceptably wrong, it doesn't matter which statistical method is applied in what way and whether the results are interpreted correctly. Secondly, the appropriateness of the statistical analysis and the correct presentation and interpretation of the results come. The presentation of the findings is also important, but statistical analysis and correct interpretation is not as important as the methodology of the study.

Table 1. Reporting guidelines for main study types (11)

Study Design	Guideline
Randomized trials	CONSORT
Observational studies	STROBE
Systematic reviews	PRISMA
Study protocols	SPIRIT
	PRISMA-P
Diagnostic/prognostic studies	STARD
	TRIPOD
Case reports	CARE
Clinical practice guidelines	AGREE
	RIGHT
Qualitative research	SRQR
	COREQ
Animal pre-clinical studies	ARRIVE
Quality improvement studies	SQUIRE
Economic evaluations	CHEERS

CONSORT: CONsolidated Standards Of Reporting Trials
STROBE: STrengthening the Reporting of OBservational studies in Epidemiology
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
PRISMA-P: Preferred Reporting Items for Systematic reviews and Meta-Analyses Protocols
STARD: STAndards for Reporting of Diagnostic Accuracy
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
CARE: CAse REports
AGREE: Appraisal of Guidelines for Research and Evaluation
RIGHT: Reporting Items for practice Guidelines in HealThcare
SRQR: Standards for Reporting Qualitative Research
COREQ: COnsolidated criteria for REporting Qualitative research
ARRIVE: Animal Research: Reporting of In Vivo Experiments
SQUIRE: Standards for QUality Improvement Reporting Excellence
CHEERS: Consolidated Health Economic Evaluation Reporting Standards

RESEARCH REGISTRY SYSTEMS

In order to prevent researchers from causing bias in the study at any stage of the research, clinical research registries, in which protocols can be uploaded, have been started to be established since 2000. It is recommended that the protocol of the study be uploaded to these registration systems before the first patient/subject is included in the study. Some journals require that the protocols of the articles sent to them be uploaded to one of the public registration systems. Journal editors and referees can evaluate whether there is deviation from the protocol at any stage of the study, which would cause bias. Protocols are commonly recorded in registries such as World Health Organization's International Clinical Trials Registry Platform (https://www.who.int/clinical-trials-registry-platform/the-ictrp-search-portal), EU Clinical Trials Register (https://www.clinicaltrialsregister.eu/), and The United States (https://clinicaltrials.gov/) to provide transparency in studies. Clinical studies conducted in our country are required to be registered on the Turkish Medicines and Medical Devices Agency's clinical research portal (https://kap.titck.gov.tr). PROSPERO (https://www.crd.york.ac.uk/prospero/) and Cochrane Collaboration (https://www.cochrane.org/) are the most

widely used registry systems for systematic review and meta-analysis studies. On the other hand, the protocols accessible from these registry systems should not be considered 100% correct.

In fact, within the scope of the concept of reproducible research, which has become popular approach in recent years, some journals have begun to request the data of the studies and the commands of the statistical methods used in the analysis as supplementary materials.

When health professionals read the articles published in literature, they should have sufficient biostatistical knowledge to understand whether the research is carried out in accordance with biostatistics principles and scientific research methods. In the following part, some basic statistical principles will be presented for this purpose.

BASIC STATISTICAL CONCEPTS

Statistics covers the methods used in planning, conducting, obtaining data, organizing, summarizing, analyzing, interpreting and reaching a decision for scientific purposes.

Population - Sample

A population is a collection of all observations/individuals/objects that have common characteristics related to the subject under study. Scientific research is not conducted on a population, but on a small subset, which is representative of that population and contains fewer observations, and is called a sample. For the sample to be representative of the population, the sample size must be sufficient and the sample should be drawn using one of the appropriate probability sampling methods. If the sampling is biased, there is no statistical method to correct this, and it is inevitable that the results will be misleading.

Parameter - Statistic

The value obtained from the sample is called statistic, and the value obtained from the population is called parameter. The main purpose of statistical methods is to estimate the unknown population parameter with the statistic calculated from the sample. Since it is studied on samples, there is always the possibility of making mistakes depending on chance, no matter what decision is made. The main purpose of complying with the principles of biostatistics in all steps of the research is to minimize this error as much as possible and to ensure that it is at the desired level.

Variable

A variable is defined as any characteristic that can take different values, such as anthropometric measures, cholesterol, hemoglobin level, and treatment duration. It's called a variable because it has values that vary from person to person or from situation to situation within the same person. Variables can be classified as dependent variables (output, explained, predicted) and independent (explanatory, predictor, exposure, covariate) variables. The dependent variable is the variable that can occur with the effect of one or more independent variables and whose relationship with the independent variables is examined. The independent variables are the variables that are thought to affect the dependent variable. The primary variable(s) of the studies should be clearly stated in the material-method section. The research design and the minimum number of samples required in the study are determined according to the primary variable(s).

Data

Data are materials collected in order to explain the study questions or to solve a problem. Data are raw materials obtained by measurements, surveys, observations, biochemical tests, biomedical imaging, etc. Data is transformed into information by processing with statistical methods. In order to achieve the targeted goals in research, it should be decided from at the beginning of the study which data will be collected, how and in which format it will be collected. It is a wrong approach to collect data that is not directly related to the study purpose in case it will be useful to us in the future. For example, adding extra questions that are not necessary to the questionnaire will reduce the reliability of the answers to other questions. If scale forms are to be used to measure the behavior/attitudes of individuals, validity and reliability studies of these scales must be carried out in our country.

Data Type

The type of data determines which descriptive statistics and statistical methods are used in the study. Data are basically divided into two groups as quantitative and qualitative (or numerical and categorical). Quantitative data can be categorized and converted into qualitative data, but it should be taken into account that there will be a loss of information in this case. While classifying the data, cut-off points should not be determined arbitrarily (to obtain statistically significant results), cut-off values should be chosen according to criteria that can be commonly accepted by everyone objectively. Arbitrarily setting cut-off points or creating new variables from variables in a dataset to obtain statistically significant results is an unethical use of statistics.

Qualitative Data

Data describing the characteristics of the individual, which do not require measurement or counting, such as gender (male-female), blood type (AB-AB-O), geographical regions, response to treatment (complete response, partial response, no response), educational status (low, medium, high), disease stages (stage I, stage II, stage III, stage IV) are called qualitative data.

Qualitative data is divided into two groups: nominal (unordered) and ordinal. Blood type, marital status, geographical area, etc. variables are nominal data, and educational level, disease stages, staging, etc. variables are ordinal data. Qualitative data can be entered into statistical software by giving numerical codes. Since the given numeric codes are not used to indicate a measured quantity, but to describe the characteristics of the subjects, arithmetic operations should not be performed on these variables. In addition, since there is no proportionality between categories in ordinal qualitative data (for example, the difference between stage I and stage II and the difference between stage III and stage IV are not equal), arithmetic operations are not performed in such data. Frequencies and percentages are used to summarize qualitative data. If only percentage values are to be given in the tables, the row or column totals should be given to know how many subjects were used in the calculation.

Advanced statistical analysis performed with qualitative data requires the use of different regression models, depending on the type of the dependent variable

(binary/dichotomous, nominal, or ordinal). While qualitative independent variables with more than two categories were included in the model; dummy variable coding is required for nominal variables. For ordinal independent variables, either a dummy variable can be coded or it can be included in the model as a numerical variable (if it is thought to be a proportional increase or decrease).

Quantitative (Numerical) Data

Data such as age, hemoglobin, leukocyte count, number of children, number of patients, and body mass index expressed as numerical values obtained as a result of measurement or counting are called quantitative data. Quantitative data is divided into discrete numerical data and continuous numerical data. Discrete numerical data are values obtained by counting and expressed as an integer number. Examples of discrete numerical data include the number of patients, the number of lymph nodes, and the number of embryos transplanted. Continuous numerical data is data that changes within a specific range and can take any value. Age, FSH, LH, hemoglobin, B12 level, etc. are examples of continuous numerical data.

When summarizing the numerical data, it should be decided which central tendency (mean, median, geometric mean, trimmed mean) and dispersion measures (standard deviation, range, interquartile range, etc.) to use according to the distribution of the data. For example, it is not appropriate to use the arithmetic mean and standard deviation in a skewed data set or in the presence of outlier/extreme values. It is also not appropriate to give the standard deviation as a measure of dispersion together with the median. Different advanced analysis methods/models are used depending on whether the quantitative dependent variables are continuous or count data, whether the relationship between the variables is linear or not, and whether parametric assumptions are satisfied or not.

All statistical methods are applied under the assumption that data are measured without error. Therefore, researchers should be sure that the data they obtained are correct, complete, valid, and reliable (1,3).

Presenting the Results with Tables and Graphs

The tables are useful tools to present the results in a compact and readily comprehensive form. When the results are given in a table, readers can easily understand the message and get an idea about the variable under study without reading the text. Constructing an appropriate table is not an easy task. It requires knowledge, skills, and experience. How the results are displayed in the table depends on what should be shown and how it should be shown to the readers. The best way to create a good table is to create several tables and choose the one that looks the best and is easy to understand. A simple table is better than a complex one. Complex tables can make it difficult to inspect and understand the information given in a table. Tables should be self-explanatory. Instead of preparing a complex table, it may be more appropriate to prepare a few simple tables.

Each table should have a title that clearly defines the contents of the table. Row and column labels must be clearly indicated. The units of measurement for the variables that appear in the rows and columns should be written as well.

The information given in the table can also be displayed graphically if desired. The main purpose is to make the information in the table clearer and easier for the reader to understand. The exact values of the results can be read from the tables, but this is not possible in the graphics. For this reason, in scientific research, it is not recommended to display the results in a graph without displaying them in a table.

The graph requires a title, a horizontal axis label, and a vertical axis label. Titles are usually written above the table and below the graph. Different graphs are used depending on the data type. Qualitative data uses bar and pie charts, while quantitative data typically uses histograms, scatter polygons, box line charts, line charts, and error bar charts.

Researchers must be careful when drawing or interpreting graphs. Changing the scale of the graphs drawn with the same data may cause different interpretations of the results. Researchers sometimes deliberately change the scale of the graphs in order to display the results as they wish. Sometimes, because their statistical knowledge level is not sufficient, they may draw the scales incorrectly (software can automatically adjust the scales) without realizing it.

The graphics' colors and resolutions should be adjusted according to their intended use. If a book or journal is to be printed in black and white, for example, shades of gray can be used instead of color graphics. To make the graphics look better in the articles, they would be drawn at 300 dpi resolution.

CONCLUSION

Researchers should be aware that research is a too difficult process to be carried out by a single person and should be conducted by a multidisciplinary team comprised of all experts who may be required. Since there is a possibility of making mistakes and biases at every stage of the research, they can only anticipate these issues and take the necessary precautions only in this manner. A biostatistics expert should join this multidisciplinary team at the beginning of the study. His knowledge will be required during the planning and executing studies, analyzing data, and presenting and reporting results. Researchers have easy access to software that can perform various statistical analyses. They should be aware, however, that if they use these software correctly and consciously, they can achieve accurate results.

Ethics Committee Approval: Since our study was a review, ethics committee approval was not required.

Conflict of Interest: None declared by the authors.

Financial Disclosure: None declared by the authors.

Acknowledgements: None declared by the authors.

Author Contributions: Idea/Concept: EK; Design: EK; Data Collection/Processing: EK; Analysis/Interpretation: EK; Literature Review: EK; Drafting/Writing: EK; Critical Review: EK.

REFERENCES

1. Sümbüloğlu V, Sümbüloğlu K. Sağlık bilimlerinde araştırma yöntemleri. 5th ed. Ankara: Hatipoğlu; 2017.
2. Esar E. The dictionary of humerous quotations. UK: Dorset Press; 1989.
3. Alpar R. Spor, sağlık ve eğitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenirlik. 6th ed. Ankara: Detay; 2020.
4. Zapf A, Huebner M, Rauch G, Kieser M. What makes a biostatistician? Stat Med. 2019;38(4):695-701.
5. Zapf A, Rauch G, Kieser M. Why do you need a biostatistician? BMC Med Res Methodol. 2020;20(1):23.
6. Altman DG. Practical statistics for medical research. London: Chapman&Hall; 1990.
7. Hofacker CF. Abuse of statistical packages: the case of the general linear model. Am J Physiol. Regulatory Integrative Comp Physiol. 1983;245(3):R299-R302.
8. Glantz SA. Biostatistics: how to detect, correct and prevent errors in the medical literature. Circulation. 1980;61(1):1-7.
9. Catalá-López F, Alonso-Arroyo A, Page MJ, Hutton B, Ridao M, Tabarés-Seisdedos R, et al. Reporting guidelines for health research: protocol for a cross-sectional analysis of the EQUATOR Network Library. BMJ Open. 2019;9(3):e022769.
10. Caulley L, Catalá-López F, Whelan J, Khoury M, Ferraro J, Cheng W, et al. Reporting guidelines of health research studies are frequently used inappropriately. J Clin Epidemiol. 2020;122:87-94.
11. equator-network.org [Internet]. Enhancing the QUAlity and Transparency Of health Research. Reporting guidelines for main study types [Cited: 2021 February 3]. Available from: https://www.equator-network.org/