OPEN ACCESS
Global Journal of Research in Medical Sciences
ISSN: 2583-3960 (Online)
Volume 03 | Issue 02 | March-April | 2023
Journal homepage: https://gjrpublication.com/gjrms/
Original Research Article

Perspective of Biostatistics in Medical Laboratory Science

[1]Dokubo, Onyeukwu Victoria, [1]Sampson Ebophni Success, [2]Nnodim Johnkennedy*

[1]Department of Medical Laboratory Science Madonna University Elele Rivers State
[2]Department of Medical Laboratory Science, Imo State University, Owerri Nigeria

Submission Date: 10 April 2023 | **Published Date:** 29 April 2023

***Corresponding author:** Nnodim Johnkennedy
Department of Medical Laboratory Science, Imo State University, Owerri Nigeria

Abstract

Although biostatistics is frequently misinterpreted in medical laboratory science, both experimenters and readers need to be familiar with it. The study of illnesses, patients, and epidemiological occurrences is made possible by biostatistical methods. The contemporary researcher cannot avoid learning about and using statistics. To design a study endeavor and to prevent egregious errors of misrepresentation, a better understanding is necessary. This paper's goal is to present a well-organized and structured point of view on the perspective of biostatistics in medical laboratory science and research, outlining the key tools for planning a scientific investigation from the formulation of a hypothesis to the presentation of the findings.

Keywords: Biostatistics, Statistics, Data Gathering, Research, Analysis Stage, Experiment and Observation.

INTRODUCTION

The use of statistical methods on biological data gathered prospectively or retrospectively is known as biostatistics. In biomedical research, biostatistics serves a crucial analytical function. Without it, it would be impossible to declare the results of any clinical trial because it serves as the foundation for drawing precise conclusions from the data gathered during a biomedical examination. According to Cadarso-Suárez and González-Manteiga (2007), "the discipline of biostatistics is nowadays a fundamental scientific component of biomedical, public health, and health services research" and that "clinical trials research, observational studies, physiology, imaging, and genomics" are some of its more established and developing areas of application.

In addition, numerous researchers have gradually drawn attention to the numerous statistical mistakes and flaws found in a large number of biomedical publications (Porter, 1999; Cooper, et al., 2002; Garca-Berthou and Alcaraz 2004; Strasak, et al., 2007; Ercan, et al., 2007; Thiese, et al., 2015). This is because biostatistics have been misused to produce several misleading results. This remark applies to "every stage of a medical research connected to data analysis; design of the experiment, data collection and pre-processing, analytic method and implementation, and interpretation," according to Ercan et al. (2007). The same data abuses are mentioned by Thiese et al. (2015), including "inaccurate application of statistical tests, lack of transparency and disclosure regarding decisions that are taken, inadequate or incorrect multivariate model development, or elimination of outliers."

The use of biostatistics in medical research begins with the planning phase of a clinical trial or lab experiment to determine the design and scale of an experiment that will ensure a good chance of finding effects of clinical or scientific relevance. To draw conclusions that are valid over a larger population, statistics is once again applied during the analysis of data (sample data). Particularly, statistics plays two functions in clinical trials and lab studies.

1. It guarantees good experimental design and efficient resource use throughout the planning phase. It is typically only feasible to conduct an adequate statistical analysis of experimental data if the design is statistically sound.

2. It is crucial for accurate results or research outcomes analysis. Experimental data-based claims must be supported by a pertinent statistical analysis. Even pilot studies and retrospective investigations typically require some statistical analysis.

Therefore, from the planning stage to monitoring, data collecting, data analysis, and result interpretation, biostatistics plays a crucial role in every stage of the research endeavor. These crucial functions that statistics perform in the realm of biomedical research were underlined by Sprent, 2003; Mandrekar and Mandrekar, 2009; & others. The scientific integrity and interpretation of the study findings in the broader medical community depend on a clear knowledge of the statistical approach as it relates to the study premise, reported results, and interpretation.

Calculating straightforward quantities like P-values, confidence intervals, standard deviations, and standard errors, or using certain common parametric or nonparametric tests, may be sufficient in straightforward scenarios. Even these straightforward ideas can occasionally be misconstrued or interpreted incorrectly by researchers from other fields who have little or no background in statistics. In order to draw meaningful conclusions from observable data, more complex research initiatives frequently require advanced statistical techniques, such as the creation and testing of mathematical models. Only those who have a thorough comprehension of both the methods' objectives and the implications of any inferences based on them should use them. Statistics offers logical ways to assess the level of uncertainty surrounding claims taken from database. At a more complex level, it offers indicators for how well data conform to a given mathematical model, for example, assesses the model's goodness of fit and, where necessary, offers estimates of specific constants or parameters in a model.

Overall, statistics is crucial to the clinical trial process at every stage, from design to conduct to interim analysis to final analysis and reporting. Statistical techniques will often be used to create the randomization schedules, provide sample size recommendations, define the standards for evaluating treatment differences, and examine response rates. To achieve a seamless integration of the statistical components into the reporting and discussion of research outcomes, close communication between statisticians—whether experts in the field or medical research personnel with a solid statistical background—and other team members is required. The Independent Data Monitoring Committee needs statistically specified data at the end of the study for final processing and interpretation.

A BIOSTATISTICS POINT OF VIEW AT THE PLANNING AND DATA COLLECTION STAGE

In order to ensure that there is little to no selection bias and that the conclusions drawn from the analysis of this data are replicable, the selection criteria used to choose the cases (i.e., data elements) to be included in the analysis must not be ad hoc. This is clear if we remember what Thiese et al. (2015) said: "Data collection and analysis mistakes directly arise from study design." The subject population, the variables to be investigated, the sufficiency of the available follow-up information for the endpoints, and characteristics of the missing data values all play a role in the case selection process, which is primarily guided by the scientific question at hand (i.e., missing by design, missing at random etc.). This is a crucial step in determining the research study's scientific value.

The next phase is to explore and define the endpoints and the explanatory (or independent or prognostic) factors in further detail after the analytic data set has been established and identified. Thankfully, the recommendations made by The STARD Statement, The STROBE Statement, and The CONSORT Statement (Bossuyt et al., 2003; Von Elm et al., 2007; and Moher et al., 2001) have helped researchers to enhance the standard of reporting study procedures and outcomes of biomedical research. The endpoint variables, or result variables, normally belong to one of three classes:

1. Categorical, which is further divided into: a. Binary or two categories (for example, limited vs. extensive stage disease; male vs. female; etc.); b. Nominal or multiple categories with no specific order (for example, blood group type: A, B, O, AB; Likert type scales: strongly agree, agree, neutral, disagree); and c. Ordinal or multiple ordered categories (for example, performance status: 0 vs. 1 (for example, overall survival, time to any recurrence etc.).

The endpoint's definition, the degree of data completeness, and the sufficiency of follow-up data are all crucial factors that aid in determining the precision and interpretability of the results in relation to the endpoint. For instance, in the case of time to event endpoints, the following issues require careful consideration: Information regarding the uniformity (and length) of follow-up (for instance, all patients are followed for a minimum of 2 years); The loss to follow-up/dropout rates (censors); and The loss to follow-up/dropout rates (censors) (for example, recurrent heart attacks in a coronary heart disease study).

Running straightforward descriptive and graphical summaries for the explanatory variables is typically advised to spot obvious deviations such outliers, scant data within particular categories, and/or suspect data points. At this point, judgments about categorizing continuous covariates and/or collapsing categories are investigated. These decisions are

often based on the following factors: the distribution of the data, the underlying biologic or clinical explanation, and the ease of interpretation.

According to Mazumdar and Glassman, statistics can be used to determine the best cut-point(s) for categorizing continuous covariates (2000). The minimum p-value strategy, two-fold cross validation, and mean are some of the strategies that derive from data and outcome management procedures.

Furthermore, as demonstrated by Altman et al. (1994), Abdolell et al., statistics is crucial in reducing inaccurate classification of continuous covariate or probable loss of information and incorrect assumption of the distribution of the data post categorization (2002). Chansky et al. (2009) showed how to categorize a continuous covariate using a data-driven technique and examined how this categorization affected the model assumptions and the results.

THE BIOSTATISTICS PERSPECTIVE AT THE ANALYSIS STAGE

Data analysis, the second and most important step in biomedical research, is important for drawing inferences from the data. At this point in the investigation, the general analytical technique typically consists of four components. Specifically: 1. Define the training and validation data sets, if applicable; 2. Clearly describe the transition from univariate to multivariate analyses, including the testing and model building framework; 3. Set the threshold for declaring statistical/clinical significance a priori for the main effects, interactions, and subgroup analyses while taking into account the multiple c.

In order to determine the prediction accuracy and practical performance of the original analysis, Harrell (2001) claims that training data sets (also known as developmental data sets) and validation data sets (also known as test data sets) are used to identify and validate statistical processes. Cross validation methods, which repeatedly divide a sizable data set into training and test sets, are two popular techniques. Another is the use of two separate data sets (with comparable data properties), one used solely for development and the other only for validation.

The nature of the data and the study hypothesis are intimately related to the biostatistical methods used to evaluate the data. For varied data treatment, biostatistical analysis makes use of terminologies and ideas including normal distribution, binomial distribution, non-parametric approaches, analysis of variance, long-tail distribution, exponential distribution, correlation, and regression. To make conclusions regarding the study's overall outcome, tests like the t-test, chi-squared test, Wilcoxon signed-rank test, Wilcoxon rank-sum test, Mann-Whitney test, and others like the log-rank test are also used.

In order to describe spread in the data field, terms like standard error, standard deviation, range, and interquartile range are frequently utilized. Many statistical analyses of measurement data, which rely on strong mathematical modeling and assumptions, are centered on the Normal distribution. The mean and standard deviation of a normal distribution define it. A measure of how effectively the sample mean predicts the population mean is the standard error, or more specifically the standard error of the mean.

A univariate analysis aids in determining the degree to which the explanatory variable independently of other variables has an impact on the outcome of interest. A summary measure is a brief description of the data that quickly communicates information about the distribution of a variable. Summary measures that are specific to a single variable are known as univariate statistics.

Analyses with only one variable can assist determine how much a covariate can affect an outcome by itself. The influence (effect size and significance) of an explanatory variable on outcome when studied alone is vulnerable to change when explored in the context of other explanatory variables, as is well known in the statistical and clinical literature. Therefore, it is critical to evaluate every potential independent influence a covariate may have on the result of a multivariate model.

The observed values of a variable are plotted on the X-axis vs the relative frequency of these values on the Y-axis in a histogram, which is a summary measure for a single variable. An illustration is the distribution of systolic blood pressures in the full data set as shown by the histogram of the systolic blood pressures of research participants.

There are numerous approaches to transition from a univariate to a multivariate model, according to Mandrekar and Mandrekar (2009). Some of them include: o Conduct a univariate analysis just on those that are clinically and/or statistically significant.

O Build a comprehensive model using all previously known significant covariates as the base and then add new covariates gradually.

O Build a multivariable model using the pool of all covariates using a selection strategy, regardless of the significance of the covariates in the univariable setting. O Explore all covariates explored in the univariable analysis in a multivariable analysis.

The number of models and covariates investigated, the sample size/number of occurrences, and the clinical relevance are often what define the statistical significance for these studies. To determine whether a variable's impact on the result varies with respect to the level of another variable, a multivariable model also involves the investigation of two-way interaction effects. Two variables are said to interact specifically if a particular combination of them produces outcomes that cannot be explained by the main effects of those variables alone. It is used to describe the relationship between two variables of interest and is also referred to as bivariate descriptive statistics. Finding potential confounding variables by analyzing the combined distribution of the exposure variable with other research variables is a crucial use of multivariate statistics.

Consider an observational study that looks at whether taking aspirin reduces the risk of myocardial infarction, as an illustration. This link may be complicated by the possibility that aspirin use is a sign of other health-related factors that affect myocardial infarction risk as well. The study's exposure variable (aspirin use) and potential confounding variables might both be described using multivariate statistics. Multivariate statistics might be as simple as tabulating means and standard deviations for continuous study variables according to aspirin use levels as aspirin use is a binary variable.

CONCLUSION

The heart and soul of biomedical research, without which there can be no meaningful results, is biostatistics. The use of statistical tools very early in the design of research is equally important for researchers. Only by utilizing the proper statistical model is it possible to estimate the sample size and determine if it has the required power or is sufficiently representative of the research population.

The adage "garbage in, garbage out" applies particularly to the use of statistics in clinical research because a study that is not statistically planned at the outset cannot produce any meaningful data, cannot be analyzed at the data analysis stage, and consequently cannot produce any useful clinical evidence. Therefore, it is crucial that all biomedical researchers not only understand statistics but also have the skills necessary to apply and evaluate the data that has been statistically analyzed.

REFERENCES

1. Abdolell M., LeBlanc M., Stephens D., et al. Binary partitioning for continuous longitudinal data: categorizing a prognostic variable. Statistics in Medicine 2002;21:3395–3409.[PubMed: 12407680]
2. Altman D. G., Lausen B, Sauerbrei W, et al. Dangers of using “optimal” cutpoints in the evaluation of prognostic factors. Journal of the National Cancer Institute 1994;86:829–835. [PubMed: 8182763]
3. Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis C.A., Glasziou, P.P., Irwig, L.M., et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003;138:W1-12. http://dx.doi.org/10.7326/0003-4819-138-1- 200301070-00010.
4. Cadarso-Suárez, C. and González-Manteiga W. (2007). Statistics In Biomedical Research; ARBOR Ciencia, Pensamiento y Cultura CLXXXIII 725 353-361 ISSN: 0210-1963
5. Chansky, K., Sculier, J.P., Crowley, J.J., et al. The International Association for the Study of Lung Cancer Staging Project: prognostic factors and pathologic TNM stage in surgically managed non-small cell lung cancer. J Thorac Oncol 2009;4(7):792–801. [PubMed: 19458556]
6. Cooper, R.J., Schriger D.L., Close R.J.H. Graphical literacy: the quality of graphs in a large- circulation journal. Ann Emerg Med. 2002;40:317–22.
7. Ercan I., Yazıcı B., Yang, Y., Özkaya G., Cangur S., Ediz, B., Kan, I., (2007); Misusage of Statistics in Medical Research. Eur J Gen Med; 4(3):128-134
8. García-Berthou E., Alcaraz C., Incongruence between test statistics and P values in medical papers. BMC Med Res Method. 2004;4:13–7.
9. Harrell, F.E., Jr. Regression modelling strategies with applications to linear models, logistic regression, and survival analysis. Springer-Verlag; New York: 2001.
10. Mandrekar, J. N. and Mandrekar, S. J., (2009). Biostatistics: A toolkit for exploration, validation and interpretation of clinical data. J Thorac Oncol.; 4(12): 1447–1449.doi:10.1097/JTO.0b013e3181c0a329.
11. Mazumdar, M., Glassman, J. R., Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. Statistics in Medicine 2000;19:113–132. [PubMed: 10623917]
12. Moher,. D., Schulz, K.F., Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. BMC Med Res Methodol 2001;1:2. http://dx.doi.org/10.1186/1471-2288-1-2.

13. Porter, A. M., Misuse of correlation and regression in three medical journals. J Roy Soc Med.1999;92:123–8.
14. Sprent, P., Statistics in medical research. Swiss Med Wkly 2003;133(3940):522–9. [PubMed:14655052]
15. Strasak, A. M., Zaman, Q., Pfeiffer, K. P., Göbel, G., Ulmer, H., (2 0 0 7). Statistical errors in medical research –a review of common pitfalls SWISS MED WKLY; 137:44–49 www.smw.ch
16. Thiese, M. S., Arnold, Z. C., Walker, S. D. (2015). The misuse and abuse of statistics in biomedical research: Biochemia Medical 2015;25(1):5–11http://dx.doi.org/10.11613/BM.2015.001.
17. Von Elm E., Altman D.G., Egger M., Pocock, S.J., Gotzsche, P.C., Vandenbroucke, J.P., The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Prev Med 2007;45:247-51. http://dx.doi.org/10.1016/j.ypmed.2007.08.012.