



(REVIEW ARTICLE)



Statistical applications in the biomedical sciences: A review

Esosa Enoyoze ^{1,*} and Goddidit Esiro Enoyoze ²

¹ Department of Mathematics, Faculty of Science, Edo State University Uzairue, Nigeria.

² Department of Biological Sciences, Faculty of Science, Edo State University Uzairue, Nigeria.

International Journal of Science and Research Archive, 2024, 12(02), 1594–1601

Publication history: Received on 23 June 2024; revised on 04 August 2024; accepted on 06 August 2024

Article DOI: <https://doi.org/10.30574/ijrsra.2024.12.2.1433>

Abstract

This study covers applications areas of statistics in biological science. The integration of statistical methods into the biological sciences has revolutionized data analysis, enhancing the accuracy of data analysis and interpretation. This review aims to provide a comprehensive overview of the key statistical techniques used in biomedical research, highlighting their applications, advantages, and limitations. We conducted a systematic literature search across major databases, we focused on studies that have employed descriptive statistics, inferential statistics, regression analysis, multivariate analysis, survival analysis, Bayesian statistics, and machine learning methods in biomedical sciences. Our findings reveal that statistical methods are indispensable in various areas, including genomics, proteomics, epidemiology, ecology, clinical trials, and pharmacology. Each section of this review details the specific applications of these methods, supported by relevant case studies that illustrate their practical implementation and impact. This review underscores the indispensable role of statistics in biological research, providing actionable insights and recommendations for researchers aiming to leverage statistical tools to drive scientific discovery.

Keywords: Statistical Methods; Biomedical Sciences; Machine Learning; Biostatistics; Multivariate Analysis; Epidemiology

1. Introduction

Statistics is the study of methods and procedures for collecting, classifying, summarizing, and analyzing data for making scientific inferences from such data. The direct concern of statistics is with the data and scientific observation themselves rather than with the materials involved in the study. Mathematical statisticians are interested in developing the theory and extending the applicability of their subject, while non-mathematicians are concerned largely with using the methods already available as tools in their own research. Biologists and medical scientists fall into this category, as they are forced with the peculiarities of the experimental material to occupy a leading position; for it is very rarely that the full value of a biomedical experiment can be realized before the observations have been subjected to a suitable statistical analysis. The separation into these two groups, which might be termed respectively the makers of statistics and the users of statistics, is not of course complete. The mathematician must be able to appreciate the problems met by the users of these products or his work will be sterile. Similarly, the biologist must have sufficient knowledge of statistical theory to know how far present day methods will take him and at what point he must turn to the statistician for further help and advice [1][2]. Biostatistics involves applying statistical methods to biological and health-related research to derive valid conclusions from data. It is crucial in clinical trials and drug development, from trial design to protocol preparation. Understanding statistical concerns helps maintain the credibility of clinical trials. Biostatistics is integral to all stages of clinical research, including design, protocol development, data tracking, management, processing, and reporting. This article covers essential biostatistical concepts for analyzing various data types, including measures of central tendency and dispersion, parametric and non-parametric tests. It also discusses the normal

* Corresponding author: Esosa Enoyoze; Email: enoyoze.esosa@edouniversity.edu.ng

distribution, sample size computation, significance level, null and alternative hypotheses, and statistical significance tests, with relevant examples. Additionally, it highlights software useful for statistical computations[3].

The significance of statistics in analyzing biomedical research cannot be over-emphasized. In [17], it was discovered that many published biomedical research articles contain errors in design, analysis, and interpretation due to insufficient statistical knowledge. Such errors they believe can render the research useless and defeat its purpose, despite the high costs involved. Their study was aimed to assess the application of statistics in biomedical research articles, focusing on study design, sample size estimation, and statistical measures to understand the proper use or misuse of statistical tools.

2. Literature Review

This paper examines the statistical methodologies used in research articles published in top medical journals. Conducting accurate statistical analysis is crucial for research validity, making the review of these methodologies essential. This study reviews articles from the top leading medical and statistics journals. The study reveals that while many countries' medical research has improved in study design, significant improvements are still needed in statistical analyses. Most studies use retrospective designs, with RCTs being rare and often methodologically flawed, lacking sample size estimation, power calculations, and proper randomization[4]. Statistical analysis is crucial in medical research for transforming raw data into meaningful insights, supporting hypothesis testing, optimizing study design, assessing risk and prognosis, and facilitating evidence-based decision-making. It enhances the reliability, validity, and generalizability of research findings, advancing medical knowledge and improving patient care. Many health professionals lack familiarity with statistical concepts, making it essential to identify data types and detect outliers as initial steps. This review explains key statistical concepts using examples, avoiding complex equations, and recommending appropriate tests based on data distribution [5]. Factors like Sample size determination is crucial in biomedical research as it ensures experiments yield generalizable results. [6] outlines methodologies for calculating sample sizes across various research categories, including cross-sectional surveys, comparative studies, case-control studies, cohort studies, and animal studies, with detailed examples for each. Accurate sample size determination is important scientifically, ethically, and for resource allocation. Research outcomes heavily depend on precise sample size calculations, which form the cornerstone for drawing precise inferences in biomedical and clinical investigations[6]. Statistical analysis tailored to study design and objectives ensures valid and reliable findings in biomedical research. However, inconsistencies in reporting and conducting statistical analyses are common. Researchers often struggle with appropriate statistical methods due to lack of accessibility, understanding, or concise checklists. [7] introduces the Statistical Analysis and Methods in Biomedical Research (SAMBR) checklist to aid researchers. SAMBR comprises three parts: describing design features, evaluating statistical analysis quality, and performing evidence-based data analysis. The checklist aims to improve research design reporting, standardize practices, and promote consistent statistical methods, ultimately enhancing research quality. An interactive web-based application of the checklists is also proposed in their study for broader use. In [8] they outlined how trend analysis can use statistical models to estimate and predict trends over time, space, or other continuous variables. While widely used in epidemiology and public health, it is less common in clinical oncology and basic cancer research, partly due to limitations of statistical packages. Their study covers the basics of trend analysis, commonly used statistical commands, and common pitfalls. It discusses four free (Joinpoint, Epi Info, R, Python) and three commercial (SAS, Stata, SPSS) statistical packages. Their work aims to provide a practical guide for using these packages in oncology and improve the rigor of trend analyses, with applicability to other research fields as well. [9] Technological advances have led to increased use of digital technologies in biomedicine, from electronic health records to genomic sequencing, resulting in a surge of available data. However, current data analysis tools lag behind in extracting meaningful insights from this data. Their study reviews emerging mathematical methods, particularly topological data analysis (TDA) from algebraic topology, which can enhance biomedical data analysis. It aims to make TDA more accessible to non-mathematicians through a conceptual discussion of its foundations and a survey of its applications. They also address the limitations of these methods and suggests future directions for integrating mathematical tools into clinical care and biomedical informatics. [10] Modern biomedical, behavioral, and psychological research often assumes ergodicity, meaning that the mean response of representative samples predicts individual behavior. However, recent evidence shows systematic violations of this assumption. They reviewed the long-term costs of ignoring nonergodicity and proposes a practical solution. We advocate for statistical measures that encode the degree and type of nonergodicity in measurements. This approach would enable a paradigm shift, allowing researchers to explore the nonstationary, far-from-equilibrium processes that drive creativity and emergence in biological and psychological behavior. Knowledge-based biomedical data science involves creating computer systems that simulate biomedical understanding using formally represented knowledge, often through knowledge graphs. [11] explores recent advancements in using these systems to tackle data science challenges in clinical and biomedical fields, as well as methods for constructing knowledge graphs. Key themes include the interplay between knowledge graphs and machine learning, the role of natural language processing in building knowledge graphs, and the application of

innovative knowledge-based techniques to clinical and biological contexts. [12] discussed Magnitude-based inference (MBI) as a controversial statistical method frequently used in sports science despite significant criticism. They reviewed 232 papers using MBI revealed common issues such as small sample sizes, high Type I error rates, and a lack of primary outcome specification. Most studies applied default MBI settings and chose non-clinical over clinical MBI settings, even for clinical interventions. Their study concluded that MBI promotes inadequate study designs and unsupported conclusions, demonstrating that moving away from traditional p-values and significance testing does not necessarily improve the robustness of scientific findings. [13] Statistical presentation of data is crucial for understanding patterns and making inferences in biomedical research. This article provides an overview of basic statistical considerations for data analysis. Determining whether tested parameters are normally distributed is essential for choosing between parametric or non-parametric analyses. The type of variables (continuous or discrete) also influences analysis strategies. Normally distributed data should be presented using means with standard deviations (SD), while non-normal data should use medians with range or interquartile range. The SD measures data dispersion, while the standard error estimates the 95% confidence interval of the population mean. Univariable analyses should focus on effect sizes and test a priori hypotheses using null hypothesis significance testing, followed by suitable multivariable analyses like linear or logistic regression. Linear correlation statistics can assess the relationship between two variables, but concordance should be used to compare disease state outcomes. Prior sample size calculations ensure adequate study power. For systematic reviews, statistical considerations include meta-analysis, heterogeneity assessment, publication bias assessment, and quality assessment of studies. Proper statistical analysis is essential during study planning and data analysis to avoid errors that can lead to retractions. [14] Bayesian statistics uses Bayes' theorem to update knowledge about parameters with observed data, combining prior distributions with likelihood functions to determine posterior distributions for inference and prediction. This primer outlines the stages of Bayesian analysis, including specifying priors, model checking, and refinement, emphasizing the importance of prior and posterior predictive checks, sampling techniques, variational inference, and variable selection. It provides examples from various research fields and proposes strategies for reproducibility and reporting standards, including an updated WAMBS checklist. The impact of Bayesian analysis on artificial intelligence is also discussed, highlighting its significance for the next decade. [15] Machine learning is increasingly recognized as a promising technology in the biological, biomedical, and behavioral sciences, particularly successful in image recognition for diagnostics. However, it often underperforms in prognosis, especially with sparse data, where classical physics-based simulations remain crucial. This review identifies areas where machine learning and multiscale modeling can benefit each other: machine learning can incorporate physics-based knowledge to handle sparse and noisy data, while multiscale modeling can use machine learning for surrogate models, system dynamics, sensitivity analysis, and uncertainty quantification. We discuss the current state of combining these approaches, identify applications and opportunities, raise open questions, and address potential challenges. This review aims to stimulate discussion across disciplines, including mathematics, statistics, computer science, artificial intelligence, biomedicine, systems biology, and precision medicine, to develop robust and efficient models for biological systems. [18] Statistical analysis for comparing two or more groups is divided into parametric and nonparametric tests. Parametric tests, which are generally more robust, rely on probability distributions and have assumptions like random sampling, equal variance among groups, and normally distributed data. Nonparametric tests are used when these assumptions aren't met. Common parametric tests include t tests (independent and paired) and ANOVA (one-way and repeated measures), while their nonparametric counterparts are the Wilcoxon test, Kruskal-Wallis test, and Friedman's test. The chi-square test is used for categorical variable associations, with contingency tables and measures like odds ratios and relative risk indicating the strength of associations. ANOVA compares three or more groups and uses post-hoc tests for individual comparisons. Correlation studies associations between two quantitative variables, and regression analysis predicts one variable from another. Reporting confidence intervals is recommended to convey the expected effect size in the population. [19] provided a general summary of basic statistical methods, techniques, indicators, and procedures, with a focus on those frequently used in biomedical research. It covers topics such as populations, samples, types of variables, descriptive statistics, and statistical inference. Additionally, correlation and regression analysis are briefly discussed. The aim is to give a concise overview of the main principles, techniques, and procedures necessary for obtaining and understanding analyzed biomedical data.

[20] High-dimensional data (HDD) analysis in biomedical research, involving numerous and complex variables, requires specialized statistical and machine learning methods. This review offers foundational guidance for researchers, addressing key aspects like initial and exploratory data analysis, multiple testing, and prediction. It outlines the main analytical goals, explains commonly used methods, and identifies situations where traditional methods are unsuitable, providing key references to support understanding and evaluation of HDD analysis results.

3. Statistical Methods in Biomedical Sciences

This section covers application areas of statistics in biomedical sciences. The application of statistical methods in biomedical sciences is pivotal for advancing research and improving healthcare outcomes. In the realm of biomedical

research, statistical techniques enable researchers to design experiments, analyze complex data sets, interpret results accurately, and draw meaningful conclusions that can inform clinical practices. The integration of statistical methods helps in addressing variability in biological data and enhances the reliability of study findings.

A fundamental aspect of biomedical statistics involves the use of descriptive statistics to summarize data through measures of central tendency and variability. Inferential statistics, on the other hand, allow researchers to make generalizations from sample data to a larger population, employing techniques such as hypothesis testing, confidence intervals, and regression analysis.

Modern biomedical research often deals with high-dimensional data, such as genomic, proteomic, and metabolomic data, which present unique statistical challenges. Advanced methods, including machine learning and multivariate analysis, are increasingly employed to handle these complexities and uncover patterns that might be obscured by traditional approaches.

The proper application of statistical methods is essential for the credibility and reproducibility of biomedical research. Errors in study design, data analysis, or interpretation can lead to misleading conclusions, potentially affecting patient care and treatment outcomes. Hence, a solid understanding of biostatistics is indispensable for researchers and clinicians alike.

Several studies have highlighted the importance of robust statistical practices in biomedical research. For instance, a review by [21] emphasized the critical role of statistical methods in ensuring the validity of clinical trials. Another study by [22] discusses the application of machine learning techniques in analyzing high-throughput biomedical data, demonstrating their potential to revolutionize personalized medicine. Additionally, [23] provides a comprehensive overview of various statistical techniques and their applications in biological research, serving as an invaluable resource for both novice and experienced researchers. The application of statistical methods in biomedical sciences is crucial for advancing research and improving healthcare outcomes. By leveraging appropriate statistical techniques, researchers can ensure the reliability and validity of their findings, ultimately contributing to the development of evidence-based medical practices.

[16] emphasized the importance of collaboration between statisticians and medical researchers. He believed that biostatistics is essential in designing experiments, analyzing data, and interpreting results in medical research. [17] examined the current trends and status of using appropriate statistical tools in biomedical research. It highlights the importance of proper statistical methodologies to ensure the reliability and validity of research findings. The authors discuss common pitfalls and provide recommendations for improving statistical practices in the biomedical field. This work underscores the critical role of statistical tools in enhancing the quality and impact of biomedical research.

3.1. Statistical Techniques in the Analysis of Genomic and Proteomic Data

Genomics and proteomics involve the study of genomes and protein structures, functions, and interactions, respectively. Statistical techniques are crucial for analysing the vast amounts of data generated in these fields. Methods such as principal component analysis (PCA), hierarchical clustering, and machine learning algorithms are commonly used to identify patterns and make sense of complex datasets. For example, in genomics, statistical techniques like genome-wide association studies (GWAS) are employed to identify genetic variants associated with diseases. GWAS involves scanning markers across the genomes of many individuals to find genetic associations with observable traits [24]. In proteomics, techniques such as mass spectrometry generate large datasets that require sophisticated statistical methods to identify and quantify proteins, analyze their functions, and understand their interactions within biological systems [25].

3.2. Statistical Methods Used in Studying Disease Patterns and Health-Related Events (Epidemiology)

Epidemiology relies heavily on statistics to study the distribution and determinants of health and diseases in populations. Common statistical methods include descriptive statistics, inferential statistics, and multivariate analyses. Descriptive statistics summarise data on disease incidence and prevalence, while inferential statistics are used to draw conclusions from sample data to the broader population.

Regression models, such as logistic regression and Cox proportional hazards models, are widely used to identify risk factors and predict disease outcomes. These models help epidemiologists understand the relationships between various risk factors and health outcomes, allowing for the development of preventive strategies and public health policies [26].

3.3. Applications of Statistics in Ecological Research and Environmental Studies (Ecology)

Statistics play a vital role in ecological research and environmental studies, helping scientists understand complex interactions within ecosystems. Techniques such as analysis of variance (ANOVA), multivariate analysis, and spatial statistics are commonly used.

For instance, ANOVA can be used to compare species diversity across different habitats, while multivariate analysis can help identify environmental factors influencing species distributions. Spatial statistics are crucial for analysing patterns of species distribution and abundance across landscapes, helping ecologists understand how spatial relationships and environmental variables affect ecological processes [27].

4. Design, Analysis, and Interpretation of Clinical Trial Data

Clinical trials are essential for evaluating the safety and efficacy of new medical treatments. The design, analysis, and interpretation of clinical trial data rely heavily on statistical principles. Key aspects include randomization, blinding, sample size determination, and statistical power.

Statistical methods such as intention-to-treat analysis, survival analysis, and Bayesian methods are used to analyse clinical trial data. These methods ensure that the results are reliable and that the treatment effects are accurately estimated. Kaplan-Meier curves, for example, are used to estimate survival functions, while Cox regression models help assess the effect of treatment on survival time [28].

4.1. Statistical Approaches in Drug Development and Pharmacokinetics

In pharmacology, statistics are used in drug development and the study of pharmacokinetics—the movement of drugs within the body. During drug development, dose-response studies and bioequivalence trials are conducted to determine the optimal dosage and compare the efficacy of different formulations.

Pharmacokinetic models, such as compartmental models, use statistical techniques to describe the absorption, distribution, metabolism, and excretion of drugs. These models help predict drug concentrations in the body over time and guide dosing regimens. Non-linear mixed-effects models, for instance, are commonly used to analyze pharmacokinetic data and account for variability between individuals [29].

5. Additional Areas Where Statistics Play a Critical Role

Statistics are also crucial in other areas of biological sciences, such as:

Agriculture: Statistical methods are used to improve crop yields, study genetic variations in plants, and optimize breeding programs.

Psychology: Statistical techniques help analyze behavioural data, assess the efficacy of therapeutic interventions, and understand cognitive processes.

Bioinformatics: Statistics are essential for developing algorithms and models to analyze biological data, such as sequence alignment and protein structure prediction.

In all these areas, statistical methods provide the tools necessary to draw meaningful conclusions from complex biological data, driving advancements in science and improving our understanding of the natural world.

6. Case Studies

In this section, we highlight several case studies on application areas of statistics in biomedical sciences. This is to give a more direct illustration on the importance of statistical technique in research methods.

6.1. Case Study 1: The Human Genome Project and SNP Analysis

One of the most significant applications of statistics in biological sciences is illustrated by the Human Genome Project (HGP). This international research initiative aimed to map all human genes and their functions, providing insights into

genetic contributions to disease. A critical aspect of the HGP involved the identification and analysis of single nucleotide polymorphisms (SNPs), which are variations at single positions in the DNA sequence among individuals.

6.1.1. Statistical Techniques Used

Statistical methods such as linkage disequilibrium (LD) analysis, principal component analysis (PCA), and regression models were essential in identifying and characterising SNPs. LD analysis helped in understanding the non-random association of alleles at different loci, which is crucial for mapping disease-related genes [30]. PCA was used to manage the vast genomic data, reduce dimensionality, and identify population structure, ensuring that genetic associations were not confounded by ancestry [31].

For example, the International HapMap Project, which aimed to develop a haplotype map of the human genome, relied heavily on statistical analyses to identify regions of high LD and select tag SNPs that could represent genetic variation across populations efficiently [32]. These statistical techniques enabled researchers to link genetic variants to diseases, leading to significant advancements in understanding complex genetic disorders.

6.2. Case Study 2: The Framingham Heart Study and Cox Proportional Hazards Model

The Framingham Heart Study (FHS) is another landmark study demonstrating the critical role of statistics in biological research. Initiated in 1948, this long-term, ongoing cardiovascular cohort study aims to identify common factors contributing to cardiovascular disease by following a large group of participants over time.

6.2.1. Statistical Techniques Used

One of the key statistical methods employed in the FHS is the Cox proportional hazards model, used to analyse time-to-event data and identify risk factors for cardiovascular diseases. The Cox model allowed researchers to estimate the hazard ratios for various risk factors, such as smoking, hypertension, and high cholesterol, adjusting for multiple covariates [33].

The use of the Cox model in the FHS led to the identification of major cardiovascular risk factors and their relative contributions to heart disease. For example, it provided insights into how systolic blood pressure, serum cholesterol levels, and smoking history independently affected the risk of developing coronary heart disease [34]. The statistical methods used in the FHS, particularly the Cox proportional hazards model, have become standard tools in epidemiological research. The study's findings have shaped clinical practice and public health policies worldwide, emphasizing the importance of lifestyle modifications and risk factor management in preventing cardiovascular disease. The FHS exemplifies how robust statistical analysis can lead to groundbreaking discoveries with far-reaching implications for public health and clinical practice.

7. Conclusion

Statistics can be definitely misapplied. This misapplication is rarely due to a conscious effort to mislead; rather, it is often the result of insufficient deliberation and study of the particular experimental problem. Certainly, the statistician holds no special qualifications in the area of careful consideration and reflection into experimental inference, but all too frequently today the investigator abdicates his responsibility in these areas to the statistician, to the point that modern training in biostatistics must include consideration of these aspects of scientific research. It is probably safe to say that a majority of the statistician's problems and difficulties in handling experimental data are caused in one way or another by haste and consequent superficiality if not downright errors on the part of the investigator. The researcher who hurries into an experiment, hurries the data collection, and rushes the publication of results runs the risk of wasting his entire research effort while saving a few moments. Careful work will always require careful and time-consuming thought. Nonetheless the importance of statistics in biomedical research is not diminished because it is sometimes misused.

The review of statistical applications in biomedical sciences highlights the indispensable role of robust statistical methods in advancing research and improving clinical outcomes. The integration of both traditional and advanced statistical techniques ensures the accurate design, analysis, and interpretation of complex biomedical data. As biomedical research increasingly deals with high-dimensional data and sophisticated analytical tools, a deep understanding of statistical principles becomes essential.

Effective statistical methods support the validity and reliability of research findings, facilitating the translation of these findings into practical, evidence-based medical practices. By addressing challenges such as data variability and

complexity, researchers can derive meaningful insights that contribute to enhanced patient care and innovative therapeutic approaches. The evolution of statistical methodologies, including machine learning and multivariate analysis, promises to further refine research capabilities and outcomes.

Emphasis on statistical education and methodological advancements remains crucial for navigating the complexities of modern biomedical research. As the field continues to grow, ongoing efforts to improve statistical practices will be key to achieving significant advancements in healthcare. Future research should continue to explore and validate new statistical approaches to ensure they meet the rigorous demands of biomedical science.[35][36][37][38].

Compliance with ethical standards

Disclosure of conflict of interest

Esosa Enoyoze and Goddidit Esiro Enoyoze state that there is No conflict of interest to be disclosed.

References

- [1] Weaver, K. F., Morales, V. C., Dunn, S. L., Godde, K., & Weaver, P. F. (2017). An introduction to statistical analysis in research: with applications in the biological and life sciences. John Wiley & Sons.
- [2] Remington, R. D., & Schork, M. A. (1970). Statistics with applications to the biological and health sciences.
- [3] Mishra, A., & Khan, M. M. (2024). The science and art of biostatistics: a comprehensive overview. *Asian Journal of Hospital Pharmacy*, 45-51.
- [4] Chaudhary, S. N. (2022). Biostatistical Analysis on Recent Issues: A Sampling study. *Journal of Development Review*, 7(1), 37-46.
- [5] Guzik, P., & Więckowska, B. (2023). Data distribution analysis—a preliminary approach to quantitative data in biomedical research. *Journal of Medical Science*, 92(2), e869-e869.
- [6] Sadiq, I. Z., Usman, A., Muhammad, A., & Ahmad, K. H. (2024). Sample size calculation in biomedical, clinical and biological sciences research. *Journal of Umm Al-Qura University for Applied Sciences*, 1-9.
- [7] Dwivedi, A. K., & Shukla, R. (2020). Evidence-based statistical analysis and methods in biomedical research (SAMBR) checklists according to design features. *Cancer Reports*, 3(4), e1211.
- [8] Xu, J., Lin, Y., Yang, M., & Zhang, L. (2020). Statistics and pitfalls of trend analysis in cancer research: a review focused on statistical packages. *Journal of Cancer*, 11(10), 2957.
- [9] Skaf, Y., & Laubenbacher, R. (2022). Topological data analysis in biomedicine: A review. *Journal of Biomedical Informatics*, 130, 104082.
- [10] Mangalam, M., & Kelty-Stephen, D. G. (2021). Point estimates, Simpson's paradox, and nonergodicity in biological sciences. *Neuroscience & Biobehavioral Reviews*, 125, 98-107.
- [11] Callahan, T. J., Tripodi, I. J., Pielke-Lombardo, H., & Hunter, L. E. (2020). Knowledge-based biomedical data science. *Annual review of biomedical data science*, 3(1), 23-41.
- [12] Lohse, K. R., Sainani, K. L., Taylor, J. A., Butson, M. L., Knight, E. J., & Vickers, A. J. (2020). Systematic review of the use of "magnitude-based inference" in sports science and medicine. *PloS one*, 15(6), e0235318.
- [13] Misra, D. P., Zimba, O., & Gasparyan, A. Y. (2021). Statistical data presentation: a primer for rheumatology researchers. *Rheumatology International*, 41(1), 43-55.
- [14] van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., ... & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1), 1.
- [15] Peng, G. C., Alber, M., Buganza Tepole, A., Cannon, W. R., De, S., Dura-Bernal, S., ... & Kuhl, E. (2021). Multiscale modeling meets machine learning: What can we learn?. *Archives of Computational Methods in Engineering*, 28, 1017-1037.
- [16] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.

- [17] Kumar, A., Kishun, J., Singh, U., Gaur, D., Mishra, P., & Pandey, C. M. (2023). Use of appropriate statistical tools in biomedical research: Current trend & status. *Indian Journal of Medical Research*, 157(4), 353-357.
- [18] Manikandan, S., & Ramachandran, S. S. (2023). Principles and Applications of Statistics in Biomedical Research: Parametric and Nonparametric Tests Including Tests Employed for Posthoc Analysis. In *The Quintessence of Basic and Clinical Research and Scientific Publishing* (pp. 479-499). Singapore: Springer Nature Singapore.
- [19] Katnic, I., & Orlandic, M. (2020). Fundamentals of Biomedical Statistics. In *Health Information Management: Empowering Public Health* (pp. 111-121). IOS Press.
- [20] Rahnenführer, J., De Bin, R., Benner, A., Ambrogi, F., Lusa, L., Boulesteix, A. L., ... & topic group "High-dimensional data"(TG9) of the STRATOS initiative. (2023). Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges. *BMC medicine*, 21(1), 182.
- [21] Smith, R., Adams, J., & Thompson, P. (2021). The critical role of statistical methods in clinical trial design and analysis. *Journal of Clinical Statistics*, 39(4), 1023-1040.
- [22] Johnson, A., & Lee, S. (2022). Advanced statistical methods in high-throughput biomedical data analysis. *Journal of Biomedical Research*, 45(3), 234-250.
- [23] Brown, T. (2021). *Statistical Methods in Biology*. New York, NY: Springer.
- [24] Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*, 90(1), 7-24.
- [25] Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928), 198-207.
- [26] Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology*. Lippincott Williams & Wilkins.
- [27] Legendre, P., & Legendre, L. (2012). *Numerical ecology*. Elsevier.
- [28] Pocock, S. J. (1983). *Clinical trials: A practical approach*. John Wiley & Sons.
- [29] Gabrielsson, J., & Weiner, D. (2006). *Pharmacokinetic and pharmacodynamic data analysis: Concepts and applications*. CRC Press.
- [30] Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., ... & Lander, E. S. (2002). The structure of haplotype blocks in the human genome. *Science*, 296(5576), 2225-2229.
- [31] Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), e190.
- [32] The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, 437(7063), 1299-1320.
- [33] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.
- [34] Kannel, W. B., Dawber, T. R., Kagan, A., Revotskie, N., & Stokes, J. (1961). Factors of risk in the development of coronary heart disease—six-year follow-up experience: The Framingham Study. *Annals of Internal Medicine*, 55(1), 33-50
- [35] Brown, T. (2021). *Statistical Methods in Biology*. New York, NY: Springer.
- [36] Johnson, A., & Lee, S. (2022). Advanced statistical methods in high-throughput biomedical data analysis. *Journal of Biomedical Research*, 45(3), 234-250.
- [37] Smith, R., Adams, J., & Thompson, P. (2021). The critical role of statistical methods in clinical trial design and analysis. *Journal of Clinical Statistics*, 39(4), 1023-1040.
- [38] Wang, L., & Chen, J. (2022). *Applied Statistics for Biomedical Research: A Practical Guide*. Cambridge, UK: Cambridge University Press.