



ISSN: 3067-0136

Academic Journal of  
Health Sciences & Research

Iris Publishers

Review Article

Copyright © All rights are reserved by Muhammad Ilyas

# The Importance and Efficacy of Advanced Statistical Software Commonly Used in Health Sciences

Muhammad Ilyas\*

PhD Nursing (Scholar), Lincoln University College, Petaling Jaya, Malaysia

ORCID ID: 0009-0006-1795-884X

**Corresponding author:** Muhammad Ilyas, PhD Nursing (Scholar), Lincoln University College, Petaling Jaya, Malaysia

**Received Date:** June 09, 2026

**Published Date:** June 26, 2026

## Abstract

**Background:** Advanced statistical software has become an essential component of modern health sciences research due to increasing complexity, volume, and diversity of biomedical data because health-related data often include a mixture of continuous, categorical, and time-dependent variables derived from clinical trials, observational studies, and large-scale databases such as electronic health records. The appropriate analysis of such data requires sophisticated statistical approaches that go beyond basic descriptive and inferential techniques. Advanced statistical software is central to make health research valid, reproducible, and clinically useful. They help choose correct study designs, avoid errors, and handle complex data structures that are now common in health sciences. They include SPSS, SAS, Stata, R (Programming language for statistics), and Microsoft Excel that play their key roles and highlighting their fundamental importance in health sciences research.

**Objective:** To evaluate and assess the importance and efficacy of advanced statistical software commonly used in health sciences research.

**Approach:** This review article systematically examines and synthesizes the existing literature on the importance and efficacy of advanced statistical software, commonly used in health sciences research.

**Significance:** The use of advanced statistical software in health sciences research is highly significant as they enhance the accuracy, reliability, and validity of research findings. These methods enable researchers to analyze complex and high-dimensional data, which are increasingly common in modern healthcare due to the growth of electronic health records, clinical trials, and large population-based studies.

**Conclusion:** Advanced statistical software is indispensable in health sciences research as they improve analytical precision, support informed decision making, and ultimately contribute to better health outcomes and more efficient healthcare systems.

**Keywords:** Statistical software; SPSS; SAS; Stata; R-programming language; Microsoft Excel

## Introduction

Advanced statistical software has become a crucial part of contemporary health sciences research because biomedical data, epidemiological studies, clinical trials, and public health studies are becoming more complex [1]. Recent advancements include next-generation statistical frameworks that enhance the examination of lifespan and survival data across a range of medical illnesses [2]. Innovations like Bayesian techniques and frailty models have improved our understanding of disease dynamics and treatment

effects, and the development of statistical methodology has been directly linked to the advancement of medical research [3].

Powerful statistical computing systems are necessary for precise data management, sophisticated modeling techniques, reproducibility, and evidence-based interpretation, all of which are critical components of modern health research. In the health sciences, software programs including IBM SPSS Statistics, R, SAS, and Stata are frequently used for descriptive statistics,



regression modeling, survival analysis, epidemiological analysis, meta-analysis, machine learning, and predictive analytics as recent literature shows these software tools greatly increase the effectiveness, validity, reproducibility, and accuracy of health-related study findings [1]. Multiple endpoints, subgroups, time-to-event data, and sizable Electronic Health Record (EHR) datasets are all part of current trials and observational studies, which call for comprehensive regression, survival, and multiplicity techniques [4].

Review Manager and Stata are in charge of systematic reviews and meta analyses, while SPSS is primarily utilized for observational and experimental approaches [5]. With extensive packages for regression, survival, and visualization, the programming language “R” offers robust, open-source support for decision analysis, health data science, cancer research, and repeatable operations [6]. The design, analysis, and interpretation of medical studies are supported by statistical principles, which guarantee validity and reliability in evidence-based healthcare decisions [7].

Furthermore, despite highlighting issues like data privacy, the combination of big data analytics and machine learning with conventional statistics is increasing capabilities in disease prediction, personalized medicine, and resource efficiency [8,9]. While survival analysis techniques manage time-to-event data crucial in clinical trials and epidemiological investigations, logistic regression is commonly employed for modeling binary outcomes [10]. Data analysis is based on descriptive statistics and inferential testing, using procedures like hypothesis formulation, assumption verification, and result reporting to guarantee rigorous scientific methodology [11]. This article will help the researchers in concluding all the advanced statistical software commonly used in health sciences research.

### Statistical Software Commonly used in Health Sciences Research

Statistical software is essential to health research because it makes sophisticated data analysis possible, increases accuracy, and promotes reproducibility. Commercial software like SPSS, SAS, and Stata are widely utilized; SPSS is the most popular for observational and experimental investigations, while Review Manager and Stata are better for systematic reviews and meta-analyses [5]. Without the need for sophisticated programming knowledge, open-source programs like R and its intuitive interface Easy R (EZR) offer accessible tools for survival analysis, competing risk analysis, meta-analysis, and sample size calculation [12,13]. In environmental epidemiology, specialized software like the Bayesian kernel machine regression (BKMR) package in R enables flexible modeling of many contemporaneous exposures [14].

Furthermore, the Sample Size & Statistical Analysis Calculator (S2SAC) and other tools improve statistical analysis and sample size estimation in well-known programs like Microsoft Excel, improving the dependability of study designs [15]. Even with the abundance of software possibilities, there are still issues with consistently disclosing software use in publications and the requirement for better software development training to guarantee high-quality

analytical code [16,17]. Commonly used statistical software in health sciences research are as under:

### Statistical Package for the Social Sciences (SPSS)

Statistical software is crucial for transforming unprocessed data into trustworthy evidence in the health sciences. For observational, experimental, and questionnaire-based research in particular, SPSS is one of the most popular tools. SPSS was the most widely used statistical tool (52.1% of publications), surpassing SAS and Stata, according to a bibliometric examination of 10,596 PubMed articles (1997, 2007, 2017). In the health sciences, SPSS is primarily associated with observational (61.1%) and experimental (65.3%) study designs [5]. SPSS is one of the most popular data analysis programs in community medicine and public health journals [18]. Clinicians, public health professionals, and students can manage data, conduct analyses, and present results properly and efficiently with SPSS because it combines strong statistical techniques with an intuitive interface.

### Importance of SPSS in Health Sciences Research

- i). SPSS is an accessible and user-friendly, having graphical user interface and menu-driven instructions make statistics and programming less difficult for students and health professionals [19,20].
- ii). SPSS has the abilities of an effective management of huge health datasets, file transformation, merging, and documentation [19, 21].
- iii). SPSS commands are integrated with advice on appropriate study design, assumptions, and critical evaluation in health statistics texts and guides [22,23].
- iv). Health sciences students can acquire critical quantitative skills for final projects and research jobs through SPSS-based instruction and workshops [24,25].

### Key Applications of SPSS in Health Sciences Research

- 1). SPSS helps in basic analyses for patient or population data include frequencies, percentages, means, and standard deviations [18,24].
- 2). SPSS helps in group comparisons i.e. chi-square test, ANOVA and t-tests to compare groups or interventions [18,24].
- 3). SPSS helps in association and prediction i.e. to investigate risk variables and outcomes, use correlation and regression analysis [20,26].
- 4). SPSS helps in coding, organizing, and evaluating patient and participants responses from surveys and questionnaires [23,24].
- 5). The most recent versions (v29) of SPSS, targeted for public health professionals and doctors, now support meta-analysis [27].

### Statistical Analysis System (SAS)

Statistical Analysis System (SAS) is an extensive software

package for data management, statistical analysis, and reporting created and maintained by the SAS Institute in 1970. From being primarily a statistical tool, it has developed into an integrated platform that facilitates enterprise-level reporting, advanced analytics, data access, transformation, and graphics [28,29]. Furthermore, SAS is a collection of preprogrammed statistical techniques integrated into a standardized software environment for data processing and analysis [28,30]. According to a bibliometric analysis of 10,596 PubMed articles from 1997, 2007, and 2017, SAS was the second most popular statistical package used in the health sciences (12.9% of articles), behind SPSS and comparable to Stata [5].

### Importance of SAS in Health Sciences Research

In health sciences research, SAS has still a vital tool, particularly for large-scale biomedical data, clinical trials, epidemiology, and survey analysis as:

- i). SAS is utilized for extensive survey and epidemiologic analyses, such as logistic regression using survey weights and replication techniques through macros like “%svy logistic regression” [31].
- ii). As a common framework for reliable treatment of missingness in clinical prediction and outcomes research, tutorials on multiple imputation for missing clinical data offer complete SAS code in addition to R and Stata [32].
- iii). Tools for personalized risk prediction from longitudinal Electronic health records (EHR) data such as the “CRcoder macro” for joint models of concurrent binary outcomes, are based on SAS/STAT and are intended to enhance customized clinical decision making [33].
- iv). With an emphasis on actual health datasets and consultative applications, recent books and reviews provide SAS and R as the main analytics platforms for health data science, including basic to advanced models e.g. GLMs, GEE, mixed models, hierarchical logistic regression which positions SAS as both a teaching and applied research workhorse [34,35].
- v). SAS, SPSS, and R were used as reference standards in a 2024 assessment of ChatGPT 4’s “Data Analyst” in epidemiology. Although it has a higher learning curve, SAS was said to provide a “comprehensive and reliable suite of analytical tools” and is still required for sophisticated techniques like Cox regression that are currently not directly executable by LLM tools. When more complicated analyses are required, the researchers specifically suggest using ChatGPT to create SAS code for local execution [20].
- vi). Recently, an SAS macro (%BBIplus) was created to calculate and display Bang’s blinding index across many randomized trial designs, increasing trialists’ access to blinding assessment [36].
- vii). Another SAS matching program reduces selection bias in observational medical research by enabling automated case-control matching on several variables with randomized

selection which is made to be used even by relatively novice SAS users [17].

viii). In order to reduce technical obstacles and increase participation in the analysis of sizable, varied EHR and survey datasets, the “all of Us” Researcher Workbench expressly incorporated SAS Studio. SAS is praised for combining a comprehensive programming language with point-and-click features, allowing users of all skill levels, and enabling initiatives on socioeconomic determinants of health and chronic illness outcomes, according to co-design with academics [37].

### Key Applications of SAS in Health Sciences Research

SAS is frequently used in epidemiology, clinical trials, and health services research for complex statistical analysis, longitudinal and survey data, and repeatable reporting as:

- 1). SAS is a benchmark biostatistical tool that is used to validate new tools like ChatGPT for descriptive, intergroup, correlation, and regression analysis in large health surveys [20,38].
- 2). SAS is second only to SPSS in a PubMed examination of 10,596 articles, it is widely used for observational and experimental designs in the health sciences [5].
- 3). To lessen selection bias, SAS macros enable multivariable matching with randomized control selection [17].
- 4). Time to event and exposure that varies over time SAS code for dividing follow-up time and reorganizing data for Cox models and causal inference as provided by tutorials [39].
- 5). PROC NLMIXED and associated algorithms are used in SAS to create nonlinear longitudinal models and proportional treatment effect models for clinical trials [40].
- 6). SAS commands and linear mixed models are used to illustrate crossover trial analysis such as carryover, period, sequence effects [41].
- 7). Along with R and Stata, SAS is considered primary software for meta-analysis as it is frequently used in conjunction with other tools for network plots or complex meta regression [42].
- 8). SAS PROC BGLIMM can be used to implement Bayesian pairwise and network meta-analysis, giving physicians an easily accessible Bayesian framework [43].
- 9). The generic SAS macro “%svy logistic regression” improves repeatability by automatically producing publication-ready tables and fitting logistic regression for complex survey data such as NHANES [31].
- 10). Analysis of electronic health records (EHRs) and surveys for precision medicine questions i.e. social determinants, multimorbidity is made possible by SAS installed on the “All of Us cloud platform” [37].
- 11). Integration with survey systems i.e. Qualtrics to automate clinical study data capture, tracking, and distribution [45].

## Stata

In health and social sciences, Stata is a popular statistical software program for data management, analysis, and graphing. Both high-level explanations of Stata's features and detailed instructions for certain techniques are included in recent literature. Stata is an all-purpose statistical program that uses a command line interface to perform analyses using typed commands for data manipulation, visualization, statistics, and repeatable reporting and it offers sophisticated programming via its Mata language and manages one dataset at a time in a spreadsheet like data editor [45].

Stata's strengths in importing a variety of data formats, performing descriptive and inferential statistics i.e. t tests, regression, survival analysis, time series, panel data models, graphics, scripting, and teamwork are highlighted in reviews of statistical software [46]. Stata is becoming more and more popular in economics and other practical subjects, including agriculture because of its effectiveness and sophisticated econometric techniques [47].

### Importance of Stata in Health Sciences Research

- i). Stata constructed 12.6% of all statistical software mentions in 10,596 PubMed publications (1997, 2007, 2017) and 38.3% of systematic reviews and meta-analyses, where it is one of the two most popular tools [5].
- ii). Stata was utilized in 46% of software-related articles in US health services research journals, making it one of the two most popular packages [16].
- iii). Geographical variation i.e. Compared to non-US work, US-authored health services publications are more likely to use Stata [16,5].
- iv). Stata is one of the most popular software programs for meta-analysis and offers thorough instructions for running fixed-effects and random-effects models as well as managing publication bias and missing data [48].
- v). In order to bridge the gap between sophisticated methodology and standard systematic reviewing, specialized Stata commands like metadta apply cutting-edge bivariate random-effects and meta-regression models for diagnostic test accuracy [47].
- vi). Stata contains a wide range of statistical tools and data management, including parameter estimation, visualizations, and biomedical kinetic analysis [49].
- vii). Stata-based texts and lectures for biomedical researchers on regression, survival, and repeated measurements [50,51].
- viii). Lasso for causal inference, meta-analysis tools, and Python integration for sophisticated health analytics are examples of modern approaches and integration [52].
- ix). Guidance comparing Stata with R/SAS/SPSS for extensive health surveys involving complex survey and epidemiologic data [53].

x). Advanced models required for clinical trials, survival analysis, and large epidemiologic studies are all the essential components of contemporary medical statistics that can be implemented with sophisticated software such as Stata [54-56].

xi). Several reviews reveal common mistakes in statistical analysis and reporting in medical publications, highlighting the fact that software including Stata is potent but needs to be used in conjunction with strong methodological knowledge and appropriate reporting guidelines [55,57,58].

### Key Applications of Stata in Health Sciences Research

- 1). According to a major PubMed-based poll, Stata is the second most used tool for systematic reviews and meta-analyses, with 38.3% of meta-analyses and systematic reviews using Stata [5].
- 2). A useful tutorial covers fixed or random-effects models, publication bias, and missing outcome data in a step-by-step meta-analysis of mental health intervention trials using Stata [48].
- 3). Meta-analysis of binomial outcomes such as prevalence and cure rates, managing extreme proportions and exact or binomial-based intervals, is made possible using a specific command called metaprop [47].
- 4). In order to compare several treatments, verify consistency, rank interventions, and evaluate publication bias, network meta-analysis (NMA) publications present complete NMA procedures in Stata [59].
- 5). Stata and SAS are the most widely used analysis tools in U.S. health care research i.e. 46% of software-related articles used Stata, particularly those written by authors in the United States [16].
- 6). Epidemiologists and public health researchers frequently utilize Stata for sophisticated survey data, regression, longitudinal or panel data, and survival analysis [60].
- 7). Numerous courses and textbooks specifically focus on applied biostatistics, data management, and repeatable workflows for health sciences professionals learning Stata [61,62].
- 8). More sophisticated data science activities in epidemiology and public health are supported by more recent Stata versions, which include tools for lasso for causal inference, extended panel data models, and Python integration [52].
- 9). List of software evaluations for quantitative bias analysis, Stata is one platform that offers tools to evaluate measurement error and misclassification in epidemiologic investigations [63].

## R (Programming language for statistics)

R (Programming language for statistics) is a free, open-source, and highly extendable statistical computing, graphics, and data research environment that is becoming more and more important in epidemiology, health, and life sciences. Its relevance in sophisticated modeling, repeatable research, and user-friendly instructional

resources is highlighted in recent literature [64,65]. R is introduced in a number of recent publications and tutorials, ranging from basic concepts to sophisticated biostatistics, epidemiology, survival analysis, and predictive analytics, frequently using actual biological datasets and step-by-step code. Research highlights R's strength and adaptability, but it also has a steep learning curve, particularly when compared to menu-driven tools [6,20].

R is described in recent literature as a sophisticated, potent, and expandable statistical programming language that supports contemporary data analysis in health sciences. It is a desirable platform for both advanced modeling and teaching due to its open-source nature, robust visuals, large package ecosystem, and repeatability tools. However, a number of publications highlight the necessity for solid instructional materials and GUIs to overcome its programming-oriented learning curve.

### Importance of R in Health Sciences Research

- i). R is extensively used by academic statisticians and is free, which is essential for academics and institutions with little funding [65,66].
- ii). R has a rich ecosystem i.e. biostatistics, machine learning, omics, clinical trials, and more are supported by tens of thousands of packages [64,67,68].
- iii). R has the features of reproducibility and transparency i.e. workflows, data, and analytics are easier to replicate and audit when using tools like R Markdown, renv, Git integration, and package-based data sharing e.g. alzverse, ADNIMERGE2 [64,69].
- iv). Free analysis and visualization for epidemiology and public health research; supports dietary indexes, sophisticated models, and TMLE [20,65].
- v). R packages for clinical and Alzheimer's research that combine trial data, documentation, and vignettes i.e. alzverse [67,70].
- vi). High-quality static and interactive images and interactive R Shiny programs for nursing and biomedical visualization [71,72].
- vii). DataSHIELD platform for federated R analysis of large, privacy-sensitive multi-site data without transferring patient-level data [73].
- viii). R has a rich training and capacity building i.e. R resources for biomedical and health data analysis, including books and toolkits [6,68].
- ix). Reproducible research while increasing rigor, dynamic documents that include text, code, and data reduce obstacles for scientists who are not programmers [69].
- x). Automation and efficiency i.e. large clinical data workflows and validation procedures are significantly accelerated by contemporary R frameworks such as tidyverse and parallel processing [72].

xi). Advanced analytics i.e. R is essential for TMLE in metabolomics visualization, microbiome pipelines, epidemiology, and machine learning-based health prediction [64,74].

### Key Applications of R in Health Sciences Research

- 1). R supports model construction, parameter estimation, and output analysis for decision-analytic models in health economics and medical decision making [75].
- 2). R is combined with Python and AI to automate data processing, validation, and analysis in clinical trials, increasing productivity and quality assurance [72].
- 3). R is utilized for data cleansing, regression, survival analysis, time series, machine learning, and medical image analysis in cancer and clinical research [76].
- 4). R is used for traditional testing, regression, time series, classification, clustering, and sophisticated visualization in epidemiology and public health analytics to enhance monitoring and decision-making, particularly in environments with limited resources [65].
- 5). R provides the foundation for interactive dashboards and visual analytics such as charts, maps, heatmaps in the fields of surveillance, interventions, policy, and education [77].
- 6). In nursing research, R Shiny makes it possible to study complex data and share discoveries through dynamic, interactive graphics [71].
- 7). End-to-end next generation sequencing (NGS) pipelines for clinical genomics and transcriptomics are supported by Bioconductor in R [78].
- 8). R is utilized in laboratory medicine for predictive modeling using regression, random forests, SVMs, clustering, and neural networks, as well as for technique comparison, precision verification, and reference interval estimation [79].
- 9). Automated biomedical analytics i.e. clinicians can use no-code solutions that combine preprocessing, statistics, survival analysis, and stacking machine learning using modular R-based pipelines [80].
- 10). R has biostatistics and fundamental techniques i.e. education and statistics for biomedical uses [3].
- 11). R has over 100 tools for systematic reviews and meta-analysis are available [67].
- 12). R has a big data and data science i.e. biomedical data science and predictive analytics [6,64].
- 13). Teaching R Shiny and Visualization i.e. educating students and clinicians on R/RShiny [81].

### Microsoft Excel

Microsoft Excel is one of the most popular tools for data management and preliminary statistical analysis in health sciences research because researchers, physicians, epidemiologists, nursing

scholars, and public health professionals find it especially useful due to its accessibility, user-friendly design, and interaction with the Microsoft Office environment [1]. It is frequently used for data entry, organization, cleaning, descriptive statistical analysis, graphical display, and initial interpretation of research findings in contemporary biomedical and healthcare investigations. Despite the growing availability of more sophisticated statistical software like R, SAS, SPSS, and Stata, recent studies show that Excel still plays a significant role in health sciences research [5].

Researchers can effectively manage massive amounts of clinical and epidemiological data with Microsoft Excel. In order to organize healthcare datasets prior to more complex statistical analysis, the software facilitates data tabulation, sorting, filtering, pivot tables, chart creation, and formula-based computations. Excel is often used by researchers in the early stages of their work to code variables, validate datasets, identify errors, and do exploratory data analysis. Literature shows that using Excel correctly may increase data accuracy, lower entry errors, and make it easier to prepare datasets for sophisticated analytical applications [82]. Microsoft Excel is frequently used in health sciences research for descriptive statistics, frequency distributions, graphical analysis, mean and standard deviation calculations, and basic inferential tests. Additionally, the program can be used to create tables, figures, and charts for presentations and research publications [1,5].

### Importance of Microsoft Excel in Health Sciences Research

i). Large volumes of clinical and research data can be gathered, stored, and managed using Excel's structured platform. Spreadsheets are used by researchers to keep track of epidemiological information, survey responses, lab results, and patient records. Data accuracy is increased and entry errors are decreased by features like data validation, filtering, sorting, and conditional formatting. Excel facilitates effective data preparation and cleansing, which are crucial stages in health sciences research [82,83].

ii). Researchers can conduct descriptive and inferential studies using Excel's built-in statistical functions, such as mean, median, standard deviation, correlation, regression, and hypothesis testing. It is especially helpful for novices and health professionals with little background in statistical programming due to its intuitive interface. Excel continues to be a popular statistical software program in health research because it makes preliminary statistical calculations and data summaries easier [82,1].

iii). Excel is used by health researchers to produce pivot tables, dashboards, graphs, and charts that enhance the communication and interpretation of research findings. Researchers can more successfully discover trends, disease patterns, and consequences when health data is represented visually [84].

iv). Clinical audits, epidemiological surveillance, and retrospective chart reviews all heavily rely on Excel because

researchers are able to monitor healthcare indicators, compute frequencies, and efficiently organize patient data. Excel macros can increase the precision and effectiveness of obtaining and processing clinical data from electronic health records [85].

v). Excel's accessibility in the majority of academic institutions and healthcare facilities is one of its biggest benefits. Excel requires less technical knowledge than more sophisticated statistical programs like R or SPSS, making it appropriate for students, medical professionals, and new researchers. Additionally, recent research has shown that Excel's accessibility and ease of use make it a useful entry-level analytical tool in healthcare settings with limited resources [86].

vi). Researchers can automate monotonous manual operations with Excel's formulae, functions, and macros. Macros increase research productivity and decrease human error by automating data cleaning, computations, and report preparation and macros make it easier to organize and tidy research data effectively [85].

vii). Excel has long been used in models for economic evaluation and health technology assessment. Excel spreadsheets are widely used by researchers for healthcare decision modeling, budget impact analysis, and cost-effectiveness analysis and even if more sophisticated programming languages like R and Python are available, Excel is still useful for more straightforward and clear health economic models [87].

viii). Excel helps researchers turn raw data into useful information for clinical and policy decision-making, which supports evidence-based healthcare. Its analytical and visualization features help medical practitioners see trends, track results, and enhance patient care [86].

### Key Applications of Microsoft Excel in Health Sciences Research

1). In health sciences, Excel is frequently used for data organization, storage, coding, and cleaning. Spreadsheets are used by researchers to organize clinical observations, survey responses, lab data, and patient demographics. Excel enhances healthcare data management's speed, accuracy, and efficiency, particularly in primary healthcare and public health settings [83,86].

2). Researchers can find missing values, duplicates, and discrepancies in datasets with the aid of Excel's filters, conditional formatting, formulae, and macros. Macros in Visual Basic for Applications (VBA) are very helpful for automating tedious data-cleaning procedures in clinical datasets and retrospective chart reviews [83].

3). Excel is widely used by researchers for regression analysis, correlation analysis, frequency distributions, descriptive statistics, and hypothesis testing. Excel is still useful for initial and intermediate statistical assessments in health sciences research, even though sophisticated statistical software might be selected for intricate analysis. Because Excel-based

models are transparent and well-known to academics and policymakers, they are still often used in health economic evaluations and health technology assessments [88].

4). Excel plays a significant role in health economics and outcomes research (HEOR), according to recent studies because its cost-effectiveness analysis, budget impact analysis, decision-analytic modeling, patient-level simulation models, and health technology assessment (HTA) are just a few of the many applications for Excel. More recent Excel features enable more flexible patient-level modeling without the need for VBA programming [88,89].

5). Excel is frequently used in clinical audits, patient outcome monitoring, cohort studies, survey data analysis, public health reporting, and disease surveillance. Because of its simplicity, it can be used for small- and medium-sized epidemiological investigations when quick data processing and reporting are necessary [83,86].

6). Researchers can visualize trends and patterns in healthcare data by creating charts, graphs, pivot tables, and dashboards using Excel. Visualization technologies support evidence-based decision-making and enhance interpretation. Research from academic health sciences libraries reveals that healthcare researchers and students are increasingly learning excellent data presenting strategies through Excel-based visualization workshops [84].

7). According to a recent research, Excel can be integrated with sophisticated analytical platforms like R and AI-based systems. Excel and statistical programming environments are being used more often by researchers to manage complex studies and increase reproducibility. Excel is still a crucial fundamental tool in sophisticated research workflows, as demonstrated by tutorials for Excel users switching to R in health economic analysis [90].

8). Large language models (LLMs) and artificial intelligence (AI) systems can automate Excel-based health economic models and technical reports, as these developments increase the effectiveness of healthcare research procedures and lessen manual labor [89].

9). Research on metabolomics and biomedicine also makes use of Excel interfaces. For instance, the IDEOM platform allows researchers to handle and interpret biochemical data in a comfortable spreadsheet environment by integrating LC-MS metabolomics analysis within Excel [91,92].

## Challenges and Limitations

Despite the significant contributions of advanced statistical software to health sciences research, several challenges and limitations continue to affect their effective utilization, interpretation, and reliability. Statistical software such as IBM SPSS Statistics, R, SAS, Stata, and Python-based analytical platforms provide sophisticated analytical capabilities; however, their misuse or inappropriate application may compromise research validity

and lead to inaccurate conclusions.

One major challenge is the limited statistical knowledge among many health sciences researchers. Advanced software can perform highly complex analyses with minimal user input, but incorrect selection of statistical tests, improper model specification, or misinterpretation of outputs can result in misleading findings. Researchers often rely heavily on automated software procedures without fully understanding the underlying statistical assumptions, which increases the risk of analytical errors and false inferences.

Recent methodological studies emphasize that inappropriate use of statistical software remains a common problem in biomedical and epidemiological research. Overall, while advanced statistical software are powerful tools in health sciences research, it is crucial to understand their limitations and approach their results with caution to ensure accurate and reliable data analysis.

Researchers and medical professionals face ongoing learning problems due to the rapid improvements in technology. New programming capabilities, interfaces, and analytical methods are frequently added to statistical software packages. It takes institutional support, interdisciplinary collaboration, and continual professional training to keep up with changing approaches. The full advantages of sophisticated statistical software in health sciences research cannot be realized without sufficient training and statistical literacy.

## Conclusion

Beyond data analysis, the importance and efficacy of advanced statistical software is crucial and effective for developing healthcare policies and making decisions. Risk prediction, illness surveillance, evidence-based clinical practice, and healthcare planning all benefit from accurate statistical modeling. High-quality systematic reviews and meta-analyses, which are crucial for developing clinical recommendations and public health initiatives, can be carried out by researchers using contemporary statistical software. As a result, statistical software expertise has emerged as a crucial skill for researchers in the health sciences and medical professionals across the globe. In health sciences research, sophisticated statistical software is essential since it provides improved analytical capabilities, accuracy, and repeatability, however, careful examination of assumptions, data quality, and researcher competence is necessary for their effective usage. To optimize their advantages and promote reliable health research, these issues must be resolved.

## References

1. Kaleeswari, M., Manikandan, G., Katerere, D. et al. (2025) A systematic review on the analysis of study design, statistical methods and statistical software tools used in South African health science research. *Discov Data* 3(1): 54.
2. Almuqrin M (2024) Next-generation statistical methodology: Advances health science research. *Alexandria Engineering Journal* 108(1): 459-475.
3. Chan B (2018) Data Analysis Using R Programming. *Advances in experimental medicine and biology* 1082 47-122.

4. Wang H, Belitskaya-Levy I, Shih M, & Lu Y (2017) *Advanced Statistical Methods* 163-179.
5. Masuadi E, Mohamud M, Almutairi M, Alsunaidi A, Alswayed A et.al. (2021) Trends in the Usage of Statistical Software and Their Associated Study Designs in Health Sciences Research: A Bibliometric Analysis. *Cureus* 13(1): e12639.
6. Musa K, Mansor W, & Hanis, T (2023) *Data Analysis in Medicine and Health using R*.
7. Martha S & Martha N (2024) The Importance of Statistics in Medical Healthcare. *International Journal for Research in Applied Science and Engineering Technology* 12(8): 303-304.
8. Rajula H, Verlato G, Manchia M, Antonucci N, & Fanos V (2020) Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment. *Medicina* 56(9): 455.
9. Yang W (2024) Research on the application of statistical methods based on big data in the medical and health field. *Applied and Computational Engineering* 87(1): 132-135.
10. Yan F, Robert M, & Li Y (2017) Statistical methods and common problems in medical or biomedical science research. *International journal of physiology pathophysiology and pharmacology* 9(5): 157-163.
11. Guetterman T (2019) Basics of statistics for primary care research. *Family Medicine and Community Health* 7(2): e000067.
12. Kanda Y (2013) Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplantation* 48(3): 452-458.
13. Shaikh M & Mubeen S (2025) Open-Source Statistical Software: Revolutionizing Health Research. *JPMA. The Journal of the Pakistan Medical Association* 75(12): 1866-1867.
14. Bobb J, Henn B, Valeri L, & Coull B (2018) Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. *Environmental Health* 17:67.
15. Nawi M (2024) Optimizing sample size estimation and statistical analysis in health sciences: A comprehensive guide using the S2SAC tool. *Medical Science* 28(149):1-16.
16. Dembe A, Partridge J, & Geist L (2011) Statistical software applications used in health services research: analysis of published studies in the U.S. *BMC Health Services Research* 11: 252-252.
17. Mortensen L, Andresen K, Burcharth J, Pommergaard H, & Rosenberg J (2019) Matching Cases and Controls Using SAS® Software. *Frontiers in Big Data* 2:4.
18. Joshi K & Jamadar D (2022) Statistical software applications and statistical methods used in community medicine and public health research studies. *National journal of community medicine* 12(03): 53-56.
19. Rahman A, & Mukhtadir M (2021) SPSS: An Imperative Quantitative Data Analysis Tool for Social Science Research. *International Journal of Research and Innovation in Social Science* 5(10): 300-302.
20. Huang Y, Wu R, He J, & Xiang Y (2024) Evaluating ChatGPT-4.0's data analytic proficiency in epidemiological studies: A comparative analysis with SAS, SPSS, and R. *Journal of Global Health* 14.
21. Bala J (2016) Contribution of SPSS in Social Sciences Research. *International Journal of Advanced Research in Computer Science* 7.
22. Abbott, M (2016) *Using Statistics in the Social and Health Sciences with SPSS and Excel*.
23. Pallant J (2020) *SPSS Survival Manual: A Step-by-Step Guide to Data Analysis Using IBM SPSS*.
24. Alili A, & Krstev D (2019) *Using Spss for Research and Data Analysis*. *Knowledge International Journal* 32(3): 363-368.
25. Handayani M, Jayadilaga Y, Fitri U, Rachman D, Istiqamah N et.al. (2023) Sosialisasi dan Pengenalan Aplikasi Pengolahan Data SPSS pada Mahasiswa Administrasi Kesehatan Fakultas Ilmu Keolahragaan dan Kesehatan. *Jurnal Informasi Pengabdian Masyarakat*.
26. Lewsey J (2015) *Medical Statistics: A Guide to Data Analysis and Critical Appraisal*. *Annals of The Royal College of Surgeons of England* 88(6): 603-603.
27. Kabir R, Syed H, Hayhoe R, Parsa A, Sivasubramanian M et.al. (2024) *Meta-analysis using SPSS: a simple guide for clinicians, public health, and allied health specialists*. *The Evidence*.
28. Hallahan C (1995) *Data Analysis Using SAS*. *Sociological Methods & Research* 23: 373-391.
29. Rodriguez R (2010) *SAS*. *Wiley Interdisciplinary Reviews: Computational Statistics* 3.
30. Thaller M (2017) *Historical Software Issue 10: Statistical Analysis System (SAS)*. *Historical Social Research* 8 88-95.
31. Muthusi J, Mwalili S, & Young P (2019) %svy\_logistic\_regression: A generic SAS macro for simple and multiple logistic regression and creating quality publication-ready tables using survey or non-survey data. *PLoS ONE* 14(9): e0214262.
32. Austin P, White I, Lee D, & Van Buuren S (2020) *Missing Data in Clinical Research: A Tutorial on Multiple Imputation*. *The Canadian Journal of Cardiology* 37(9): 1322-1331.
33. McAvay G, Murphy T, Agogo G, & Allore H (2019) *CRcoder: An Interactive Web Application and SAS Macro to Support Personalized Clinical Decisions*. *The Permanente journal* 24(1).
34. Wilson J, Chen D, & Peace K (2023) *Statistical Analytics for Health Data Science with SAS and R*.
35. Rahnvard A (2023) *Statistical Analytics for Health Data Science with SAS and R*. *Journal of the American Statistical Association* 119 786-787.
36. Qin Z, Yu Y, Gu H, Shi D, Wang Z et.al. (2024) SAS macro program for Bang's Blinding Index to assess and visualize the success of blinding in randomized controlled trials. *General Psychiatry* 37(6): e101578.
37. Humes I, Shyr C, Dillon M, Liu Z, Peterson J et.al. (2024) Empowering the biomedical research community: Innovative SAS deployment on the All of Us Researcher Workbench. *JAMIA* 31(12): 2994-3000.
38. Ruta M, Gaidici T, Irwin C, & Lifshitz J (2025) ChatGPT for Univariate Statistics: Validation of AI-Assisted Data Analysis in Healthcare Research. *Journal of Medical Internet Research* 27:e63550.
39. Iwagami M, Ishimaru M, Takeuchi Y, & Shinozaki T (2025) Methodological Tutorial Series for Epidemiological Studies: When and How to Split the Follow-up Time in the Analysis of Epidemiological or Clinical Studies with Follow-ups. *Journal of Epidemiology* 35(4): 161-169.
40. Wang G, Wang W, Mangal B, Liao Y, Schneider L et.al. (2024) Novel non-linear models for clinical trial analysis with longitudinal data: A tutorial using SAS for both frequentist and Bayesian methods. *Statistics in Medicine* 43(15): 2987-3004.
41. Lim C & In J (2021) Considerations for crossover design in clinical study. *Korean Journal of Anesthesiology* 74 293-299.
42. Tantry T, Karanth H, Shetty P, & Kadam D (2021) Self-learning software tools for data analysis in meta-analysis. *Korean Journal of Anesthesiology* 74(5): 459-461.
43. Rott K, Lin L, Hodges J, Siegel L, Shi A et.al. (2021) Bayesian meta-analysis using SAS PROC BGLIMM. *Research Synthesis Methods* 12(6): 692-700.
44. Parsons F, Mota S, & Quan Y (2018) *Managing Qualtrics Survey Distributions and Response Data with SAS*.

45. Dubrovin V, Deineha L, & Yatsenko A (2023) Statistical analysis software. *Electrical Engineering and Power Engineering*.
46. Bhardwaj R & Kaushik D (2024) Comparative analysis of statistical software usage in agricultural research: A review. *International Journal of Advanced Biochemistry Research* 8(3S):727-728.
47. Nyaga V & Arbyn M (2022) Correction: Metadata: a Stata command for meta-analysis and meta-regression of diagnostic test accuracy data – a tutorial. *Archives of Public Health* 80:216.
48. Chaimani A, Mavridis D, & Salanti G (2014) A hands-on practical tutorial on performing meta-analysis with Stata. *Evidence-Based Mental Health* 17(4): 111-116.
49. Boston R, & Sumner A (2003) STATA: a statistical analysis system for examining biomedical data. *Advances in experimental medicine and biology* 537: 353-369.
50. Dupont W (2009) *Statistical Modeling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data*.
51. Weinberg S, & Abramowitz S (2020) *Statistics Using Stata*.
52. Ho A, Huynh K, Jacho-Chávez D, & Rojas-Baez D (2021) Data Science in Stata 16: Frames, Lasso, and Python Integration. *J. Stat. Softw* 98.
53. So, H, Oz U, Griffith L, Kirkland S, Raina P et.al. (2020) Modelling Complex Survey Data Using R, SAS, SPSS and Stata: A Comparison Using CLSA Datasets. arXiv: Computation.
54. Chan W (2013) *Statistical Methods in Medical Research. Model Assisted Statistics and Applications* 8 83-84.
55. Nieminen P (2020) Applications of Medical Informatics and Data Analysis Methods. *Applied Sciences* 10(20): 7359.
56. Yardibi F, Chen C, Aladag Ç, & Kose O (2025) Current Trends and Future Directions of Statistical Methods in Medical Research: A Scientometric Analysis. *Journal of evaluation in clinical practice* 31(6): e70257.
57. Dwivedi A (2022) How to write statistical analysis section in medical research. *Journal of Investigative Medicine* 70(8): 1759-1770.
58. Mansournia M, & Nazempour M (2024) Recommendations for accurate reporting in medical research statistics. *The Lancet* 403(10427): 611-612.
59. Shim S, Yoon B, Shin I & Bae J (2017) Network meta-analysis: application and practice using Stata. *Epidemiology and Health* 39: e2017047.
60. Gutierrez R.G (2010) Stata. *WIREs Comp Stat* 2(6): 728-733.
61. Unguryanu T, & Grjibovski A (2014) Introduction to stata - software for statistical data analysis 21(1): 60-63.
62. Goodman M (2017) *Biostatistics for Clinical and Public Health Research*.
63. Gerlach-Wood C, Tilling K, Bartlett J, & Hughes R (2025) Quantitative bias analysis for mismeasured variables in health research: a review of software tools. *BMC Medical Research Methodology* 25(1):187.
64. Giorgi F, Ceraolo C, & Mercatelli D (2022) The R Language: An Engine for Bioinformatics and Data Science. *Life* 12(5): 648.
65. Daswito R, Besral B, & Ilmaskal R (2023) Analysis Using R Software: A Big Opportunity for Epidemiology and Public Health Data Analysis. *Journal of Health Sciences and Epidemiology* 1(1):1-5.
66. Morton C, Devito N, Morley J, Dillingham I, Schultze A et.al. (2022) Software development skills for health data researchers. *BMJ Health & Care Informatics* 29(1): e100488.
67. Lortie C (2022) Doing Meta-Analysis with R - A Hands-On Guide. *J. Stat. Softw* 102.
68. Dinov I (2023) *Data Science and Predictive Analytics: Biomedical and Health Applications using R. Data Science and Predictive Analytics*.
69. Siraji M, & Rahman M (2023) Primer on Reproducible Research in R: Enhancing Transparency and Scientific Rigor. *Clocks & Sleep* 6(1): 1-10.
70. Donohue M, Hussen K, Langford O, Gallardo R, Jimenez-Maggiara G et.al. (2026) Alzheimer's Clinical Research Data via R Packages: the alzverse. *Alzheimers Dement* 22(2): e71152.
71. Heinsberg L, Koleck T, Ray M, Weeks D, & Conley Y (2022) Advancing Nursing Research Through Interactive Data Visualization with R Shiny. *Biological Research for Nursing* 25(1): 107-116.
72. Ravula R (2025) Automation in Statistical Programming: Advancing Clinical Research Through R, Python, and AI Integration. *European Journal of Computer Science and Information Technology* 13(9): 127-140.
73. Marcon Y, Bishop T, Avraam D, Escribà-Montagut X, Ryser-Welch P et.al. (2021) Orchestrating privacy-protected big data analyses of data from different resources with R and DataSHIELD. *PLoS Computational Biology* 17(3): e1008880.
74. Smith M, Phillips R, Luque-Fernández M, & Maringe C (2023) Application of targeted maximum likelihood estimation in public health and epidemiological studies: a systematic review. *Annals of epidemiology* 86:34-48. e28.
75. Jalal H, Pechlivanoglou P, Krijkamp E, Alarid-Escudero F, Enns E et.al. (2017) An Overview of R in Health Decision Sciences. *Medical Decision Making* 37(7): 735-746.
76. Gandhi M, Tripathy S, Pawale S, & Bhawalkar J (2024) A narrative review with a step-by-step guide to R software for clinicians: Navigating medical data analysis in cancer research. *Cancer Research, Statistics and Treatment* 7(1): 91-99.
77. Ofori M, Lartey S, Durneva P, Jha N, Mittal N et.al. (2025) Visual communication of public health data: a scoping review. *Frontiers in Digital Health* 7:1555231.
78. Sepulveda J (2020) Using R and Bioconductor in Clinical Genomics and Transcriptomics. *JMD* 22(1): 3-20.
79. Ma C & Qiu L (2025) Unveiling the power of R: a comprehensive perspective for laboratory medicine data analysis. *Clinical Chemistry and Laboratory Medicine (CCLM)* 63(8):1458-1471.
80. Miller H, Goodin D, & Frieboes H (2026) An automated software methodology for biomedical statistics, data pre-processing, and machine learning. *Computer methods and programs in biomedicine* 273: 109096.
81. Nordmann E, McAleer P, Toivo W, Paterson H & DeBruine L (2022) Data Visualization Using R for Researchers Who Do Not Use R. *Advances in Methods and Practices in Psychological Science* 5(2): 251524592210746.
82. Elliott A. C, Hynan L. S, Reisch J. S, & Smith J. P (2006) Preparing Data for Analysis Using Microsoft Excel. *Journal of Investigative Medicine* 54(6): 334-341.
83. Bauzon J, Murphy C, & Wahi-Gururaj S (2021) Using macros in microsoft excel to facilitate cleaning of research data. *Journal of Community Hospital Internal Medicine Perspectives* 11(5): 653-657.
84. LaPolla F.W.Z (2020) Excel for data visualization in academic health sciences libraries: a qualitative case study. *Journal of the Medical Library Association* 108(1): 67-75.
85. Bauzon J, Romero-Velez G, Sehnem L, Shin J, Siperstein A, Jin J (2025) Comparative Analysis of the Accuracy of Microsoft Excel Macros in Retrospective Chart Review Studies. *Journal of Surgical Research* 311: 92-97.
86. Yeni A.C, Syahputra M, Hanifa A, Werleam I, Fauzi D (2025) *Journal of Artificial Intelligence and Digital Business (RIGGS)* 4 Nomor 3.
87. Incerti D, Thom H, Baio G, Jansen J.P (2019) R You Still Using Excel? The Advantages of Modern Software Tools for Health Technology Assessment. *Value Health* 22(5): 575-579.
88. Blissett R.S, SullivanInola W, Adam S & Nagy I (2025) Efficient patient-level health economic modelling in Excel without VBA: A Tutorial. medRxiv.

89. Rawlinson W, Teitsson S, Reason T, Malcolm B, Gimblett A et.al. (2025) Generative Artificial Intelligence to Automate the Adaptation of Excel Health Economic Models and Word Technical Reports. *Value in Health* 28(11): 1683-1689.
90. Naylor N.R, Williams J, Green N, Lamrock F & Briggs A (2023) Extensions of Health Economic Evaluations in R for Microsoft ExcelUsers: A Tutorial for Incorporating Heterogeneity and ConductingValue of Information Analyses. *Pharmaco Economics* 41(1): 21-32.
91. Creek D.J, Jankevics A, Burgess K.E.V, Breitling R & Barrett M.P (2012) IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data. *Bioinformatics* 28(7): 1048-1049.
92. Nyaga V, Arbyn M, & Aerts M (2014) Metaprop: a Stata command to perform meta-analysis of binomial data. *Archives of Public Health* 72(1): 39.