Interpreting Pharmacological Data: Statistical Considerations in R Studio

Nadia Idrees[1], Rabia Munir[1], Muhammad Mubashir Khan[1], Shamila Majeed[1], Sidra Altaf[2,*] and Tasawar Iqbal[3]

[1]Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistan
[2]Department of Pharmacy, University of Agriculture, Faisalabad, Pakistan
[3]Institute of Physiology and Pharmacology, University of Agriculture Faisalabad, Pakistan
*Corresponding author: sidra.altaf@uaf.edu.pk

Abstract

Statistical analysis is a major component of pharmacological research, as it allows researchers to make sense of complex data and draw appropriate conclusions. This article presented key statistical considerations in the context of analyzing pharmacological data using the R and R Studio, focusing on the relevance of R in modern pharmacological research. Our purpose here was to present non-exhaustive methods of statistical analysis, including descriptive statistics, inferential statistics, regression, and advanced modeling in the exploration of pharmacology. We focused on the management of a variety of pharmacological data, including dose-response, efficacy, toxicity, and time-course data and dealing with complexities such as missing data, outlier considerations, and the need for normalization. The paper also covers basic strategies for modeling pharmacokinetics (PK) and pharmacodynamics (PD) - compartmental and non-compartmental analysis, non-linear regression, and survival analysis. It discusses well-established approaches to reproducibility, selecting doable statistical tests, and avoiding issues such as 'overstatement' and 'p-hacking.' Finally, the paper looks at the future of statistical pharmacology in R, where there are new tools and approaches - machine learning and Bayesian analysis - anticipated will be more important in the future. In summary, our goal is to help researchers develop an entire walk-through of their statistical pharmacology projects in R, producing reliable and reproducible results.

Keywords: pharmacological data, Statistical analysis, R Studio, Regression modeling, Pharmacokinetics, Reproducibility

Cite this Article as: Idrees N, Munir R, Khan MM, Majeed S, Altaf S and Iqbal T, 2025. Interpreting pharmacological data: Statistical considerations in R studio. In: Abbas RZ, Akhtar T and Arshad J (eds), One Health in a Changing World: Climate, Disease, Policy, and Innovation. Unique Scientific Publishers, Faisalabad, Pakistan, pp: 390-397. https://doi.org/10.47278/book.HH/2025.470

A Publication of Unique Scientific Publishers

Chapter No: 25-055

Received: 12-Feb-2025
Revised: 21-Apr-2025
Accepted: 13-May-2025

Introduction

Pharmacological data analysis is a critical aspect of the drug discovery, development and clinical processes. This step is what enables us to clearly understand the outcomes of experiments, evaluate the potential for therapeutics of compounds, and analyze the safety efficacy of a given compound. A rational and systematic approach to data analysis allows for an understanding of the compound, which facilitates informed decisions early in the drug design process and continues through post market surveillance. This also aids in optimizing dosing strategies and complying with regulatory requirements leading to improved patient outcomes (Zhao et al., 2022). In pharmacology, statistics are very important in that they provide Researchers with a means of making objective assessments of biological responses, and quantifying variability in the data. such as being able to measure the relationship of the dose-response, determining the bioequivalence of formulations of drugs, judging the safety, or efficacy of medical treatments. The application of statistics means that researchers can separate real effects from random variation allowing them to properly observe and describe study data and draw scientifically valid conclusions (Michel et al., 2020). R Studio is an integrated development environment (IDE) that commonly provides researchers with the ability to perform statistical analysis in pharmacology using the R programming language. R Studio helps researchers with powerful data management, access to large number of statistical packages, and robust workflows for reproducible research. R Studio allows researchers to work with large datasets, sophisticated statistical models, and output graphs and reports of high quality. The value of improved transparency and efficiency has made R Studio a tool widely embraced for data analysis in pharmacology (Okoye & Hosseini, 2024).

2. Data Preparation in R Studio

2.1. Importing and Cleaning Pharmacological Datasets

In pharmacology research, careful data preparation is necessary to verify that the datasets are suitable for statistical analysis and identifying key findings. R Studio offers a variety of tools and functions to assist the researcher with importing and cleaning pharmacological data. One of the most frequently utilized functions to import comma-separated values (CSV) files is read.csv(), as CSV is a common file type used to store experimental data. In addition, researchers can use the read_excel() function from the readxl package to work with Excel files, which provide additional flexibility to work with data. This can include options to select certain sheets or ranges of cells, which is particularly useful for complex datasets. Once the data has been imported, it may require cleaning to address issues such as missing or erroneous values, and making sure that the variable types are formatted correctly. The tidyverse packages, especially dplyr and tidyr, help

make data cleaning efficient and prepare data in a format that can create meaningful analyses (Fu et al., 2021). The dplyr package can make it easy for researchers to filter data, select variables, and to create new columns using a clear and understandable format to syntax. Functions such as mutate() can be useful for making derived variables, while functions such as filter() and select() can be used to analyze only necessary portions of the data (Kabacoff, 2024).

2.2. Handling Missing or Outlier Data

Dealing with missing or outlier data is an important step when preparing pharmacological datasets for analysis. There are several tools to successfully navigate these challenges. For example, the na.omit() function deletes entire rows while retaining complete cases of missing values before the analysis. This function is useful when incomplete data would distort statistical results. However, whenever deletion of data may cause the loss of meaningful information, imputation is used. Imputation helps to estimate missing values on the basis of data trends that already exist (Zhang & Thorburn, 2022). Outliers generally skew results and negatively impact the accuracy of models and can typically be detected based on statistical summaries or visualizations. The boxplot.stats() function in R does a great job of finding outliers based on what values fall above and below the interquartile range. After finding outliers, the filter() function in the dplyr package can be used to keep or remove values using your discretion (Mahoney, 2021).

2.3. Descriptive Statistics

Descriptive statistics provide the basic foundation for understanding pharmacological data, supporting the researcher with the central tendencies, variability, and distribution of the variables. In R Studio, the summary() function is a fast way to summaries of summary statistics like minimum, maximum, mean, median and quartiles for each variable in the data frame. This function is useful at the start of data exploration in an R script, allowing the scientist to see if there are obvious anomalies or unexpected trends emerging in their dataset. Variability is most common assessed with the sd() function, which provides the standard deviation as a measure of how different individual observations are from the mean of the variable. A lower standard deviation shows more uniformity across observations, which can be very desirable quality in pharmacological research (Bulanov et al., 2021).

3. Exploratory Data Analysis (EDA)

3.1. Visualizing Pharmacological Data

Exploratory Data Analysis (EDA) is essential in pharmacological studies because it enables researchers to visually explore any patterns, trends, and outliers in datasets. R studio, with the GGPLOT2 package, provides a powerful solution for producing visual outputs that are clean, clear, and customizable. Scatter plots are often used to assess the relationship between two continuous variables, such as drug dose and level of response, allowing researchers to see trends or possible non-linear relationships. These visualization methods can uncover important patterning in the data and enhance the communication of findings to interested stakeholders; all of this reinforces the importance of EDA in data analysis (Bao et al., 2024).

3.2. Identifying Patterns and Distributions

Identifying patterns and assessing variable distributions are fundamental parts of data analysis in pharmacology. Histograms are commonly used to explore the distribution frequency of continuous variables (e.g., a drug concentration or a patients' level of response). By separating the continuous data into intervals (bins) and plotting the number of observations in each bin, histograms provide clues on data skew, modality, and outliers (Hedaya, 2023). The benefit of this type of visualization especially applies to determining whether or not the data meet the assumptions of normality needed for certain statistical testIn essence, density plots show a smoothed representation of histograms, enabling you to visualize the shape of the distribution more directly, and avoid the issue of being subject to arbitrary bin widths of histogram charts (Shatz, 2024).

3.3. Group Comparisons

Group comparisons are an important element of pharmacology research to help inform whether different treatments or dosages yield different effects. Visual inspection can often act as a useful first step toward informing group differences before conducting formal statistical tests. In R Studio, researchers can create visualizations like box plots, bar charts and violin plots to visually compare variable responses across an array of treatment groups or dosage levels (Kim et al., 2020). These representational methods are very useful for identifying shift in the central location, spread, and outliers, thus offering great value in helping to understand the reliability and consequence of treatments (Schwartz et al., 2020).

4. Inferential Statistical Techniques

4.1 Hypothesis testing

Methods of inference are necessary in pharmacology to make population-wide conclusions based on sample data. Hypothesis testing allows researchers to determine if observed differences or relationships in pharmacology datasets are statistically meaningful. Many built-in functions are available in R Studio to conduct this analysis. For example, the function t. test () can be used to compare the mean of a sample to a population mean, or the means of two independent or paired groups (e.g. control vs 'treatment' groups) to examine whether a drug has a significant treatment effect (Ohlsson & Kendler, 2020).To assess more than 2 groups, you can use the aov() command/function in R. ANOVA (Analysis of Variance) is a good way to determine dose-response relationships or examine the effects of different formulations of a drug. ANOVA simply tests if there is a statistically significant difference between the means for the groups while controlling for variability in the data (Elkomy et al., 2024).

4.2. Correlation and Regression Analysis

Correlation and regression analyses are critical components of in pharmacology for exploring relationships of variables. Correlation measures both the strength and direction of the association between two variables that are both continuous. A typical function in R Studio to get a Pearson or Spearman correlation coefficient is cor(). Pearson correlation is useful for normally distributed data while Spearman would be useful for non-parametric data, or where the relationship is monotonic but not necessarily linear. These methods can be utilized to assess if two variables, for example, a drug dosage and patient response, are related and how they are related. Regression analysis extends from correlation and looks at the relationship in greater detail (Olsen et al., 2020). Linear regression is easily performed with the lm () function in R, in which researchers can estimate the effect of an independent variable (for example - drug dosage) on a dependent variable (e.g. response) providing information about direction, strength, magnitude of the effect and predictions (Jarantow et al., 2023).

4.3. Non-parametric tests

Non-parametric tests can be very helpful in pharmacological projects whenever their data do not meet the assumptions of normality or the number of observations is small. Non-parametric methodologies provide valid alternatives to parametric methods because the data are being evaluated at the level of rank rather than at the actual value. In R Studio, the wilcox.test() function, to run the Wilcoxon Rank-Sum (Mann-Whitney U) test is designed to compare two independent groups. The Wilcoxon Rank-Sum test is useful for assessing treatment effects when the response variable does not follow a normal distribution (Dagenais et al., 2022).

5. Pharmacological Modeling in R

5.1. Dose-response Analysis

Evaluating the response to a drug dose is a basic element of pharmacological research, in that responses allow the study of the relationship between the administered dose and the biological effect. In study R, sigmoidal EMAX models are often used to describe that relationship when responses increase in amounts with increasing doses to a saturation point. The DRC package (Dose-Response Curves), and NPLR (Non-Parametric Logistic Regression), are a few methods they use to develop and analyze their predictive models (Coleman, 2020).

The DRC package contains numerous dose-response models including the four-parameter logistic model (4PL), which is the most common model used in Sigmoidal EMAX modeling. In this model, the values that are estimated include (and can be interpreted as) the maximum effect (EMAX, the point at which a drug has its greatest effect), the ED50 (the dose needed to achieve 50% of the maximum effect), and the slope of the response curve (or how quickly a drug is eliciting its effect) (Gerard, 2021).

5.2. PK/PD Modeling Basics

Pharmacokinetic (PK) modeling and pharmacodynamic (PD) modeling are distinct, although inherently linked, to understand how a drug moves through the body and how that drug distributes across biological systems. PK studies most commonly use one-compartment and two-compartment models. A one-compartment model assumes that the drug is evenly distributed throughout the body; consequently, its concentration would decline uniformly over time. A two-compartment model differentiates a central compartment (i.e., the blood) and peripheral (i.e., body) compartments. It provides a more advanced representation of drug distribution and allows modeling how the drug is removed from the central circulatory system (Zou et al., 2020).

5.3. Bioequivalence testing

The bioequivalence testing process is important in determining if two formulations of a drug, usually a generic drug and a reference product, have the same pharmacokinetic properties and therefore can be assumed to be equally safe and effective. In R, bioequivalence testing often begins with the Two One-Sided Tests (TOST) approach. In TOST, you need the ratio of a pharmacokinetic parameter, such as Cmax or AUC, to fall in some defined equivalence range (for example, 0.8 to 1.25) between the two formulations. The TOSTER package allows users to implement the TOST test in R to test the two one-sided hypotheses that are required for bioequivalence (Charoo, 2020)(Miranda et al., 2022).

6. Advanced Statistical Techniques

6.1. Mixed-effects Models for Longitudinal Pharmacological Data

Mixed effects models are a useful way to analyze longitudinal pharmacology data, particularly if you have repeated measures on subjects with multiple measurements taken over time. Mixed effects models have fixed effects (i.e., drug doses, treatment group) and random effects (i.e., non-repeatable subject-to-subject variability), which is particularly well suited for data that involves multiple observations per subject. In R Studio, for example, you could take advantage of lme4 or nlme(R functions lmer() and lme() to fit mixed effects models in an efficient manner. (Jordan et al., 2020).Most researchers rely on the lmer( ) function from the lme4 package to fit linear mixed models, as it has the ability to fit both random intercepts as well as slopes across individuals making it a convenient way to assess how different individuals vary in their treatment responses over time between animals or subjects (Brown, 2021).

6.2. Bayesian Approaches

Bayesian methods represent a flexible and useful way to analyze pharmacological outcomes, as they allow for the incorporation of prior knowledge and the potential to update prior knowledge as new information (and data) become available. In R Studio, RStanArm and brms are two packages that can be utilized to implement Bayesian methods, but both require Stan to run in the background. Both package allow the gradual construction of complex models that researchers can condition on using Markov Chain Monte Carlo (MCMC) methods which thus allow researchers to formulate rich posterior distributions (Ruberg et al., 2023).

6.3. Machine Learning for Pharmacology

Machine learning has been increasingly adopted in pharmacology to improve predictive modeling and aid in data-driven decisions. Within R Studio, there are popular packages to create and evaluate machine learning models, such as randomForest, caret, and xgboost. The caret package provides a consistent and uniform interface for training, tuning, and evaluating many different types of classification and regression algorithms, vastly simplifying the work needed to develop a model (Ota & Yamashita, 2022).

RandomForest is a great approach to analyzing voluminous datasets with large numbers of predictors, and it is particularly helpful for classification problems like predicting drug responses and looking for possible adverse events. XGBoost is an advanced (state-of-the-art) boosting algorithm, particularly in terms of predictive accuracy. This technique is very relevant to pharmacology, and represents a way to work with complex, high-dimensional data. Both machine learning methods offer tremendous value in terms of identifying meaningful patterns, assisting with treatment response predictions, and representing a pathway to shorten, and speed up, the drug development process (Obaido et al., 2024). Table 1 provides interpretation of pharmacological data using R Studio, along with tools, purposes and considerations at each stage.

Table 1: Interpretation of pharmacological data using R Studio, along with tools, purposes and considerations at each stage

Step	Tool/Package	Purpose	Statistical Method	Considerations	Additional Notes	References
Data Import	read.csv(), readxl::read_excel()	Importing data from CSV or Excel files	Import data	Data formatting	Check for correct data types and columns	(Monkman, 2024)
Data Cleaning	na.omit(), impute(), dplyr::filter()	Cleaning and preprocessing data	Clean data	Handling missing data appropriately	Consider imputation methods or data deletion strategies	(Kuhn & Silge, 2022)
Handling Missing Data	boxplot.stats()	Dealing with missing values or outliers	Imputation	Detecting outliers and anomalies	Use boxplots or histograms for outlier detection	(Iqbal, Altaf, Salma, et al., 2024)
Descriptive Statistics	summary(), sd(), var()	Summarizing key statistics like mean, median, SD, variance	Mean, Median, SD	Understanding distribution of data	Report variability and central tendency	(Alabi & Bukola, 2023)
Exploratory Data Analysis	ggplot2	Exploratory analysis of data to find patterns	Visualization	Choosing appropriate visualizations	Choose plot types based on data structure	(Altaf et al., 2024)
Visualizing Data	ggplot2	Creating visual representations of data	Scatter plots, Box plots	Clear labeling and axis scaling	Avoid overfitting	(Iqbal, Altaf, Basit, et al., 2024)
Group Comparisons	ggplot2	Comparing different groups visually	Group comparison	Proper group comparison	Group stratification and factor interaction	(Myint et al., 2020)
Hypothesis Testing	t.test, aov (Iqbal, Altaf, Salma, et al., 2024), chisq.test	Testing hypotheses for significance	t-test, ANOVA, Chi-square	Assumption checks for parametric tests	Use multiple tests if necessary	(Iqbal et al., 2023)
Correlation Analysis	cor()	Assessing relationships between variables	Pearson/Spearman Correlation	Data normalization and transformation	Check assumptions like linearity and homoscedasticity	(Altaf et al., 2023)
Regression Analysis	lm()	Predicting values using linear or multiple regression models	Linear Regression	Model diagnostics	Evaluate residuals and multicollinearity	(Bayman & Dexter, 2021)
Non-parametric Tests	wilcox.test, kruskal.test	Performing tests that do not rely on parametric assumptions	Wilcoxon, Kruskal-Wallis	Choosing appropriate non-parametric tests	Select tests that work for non-normally distributed data	(Altaf & Iqbal, 2023)
Bayesian Inference	rstanarm, brms	Incorporating prior knowledge in statistical models	Bayesian Inference	Selecting proper priors and distributions	Understand model uncertainty and posterior distributions	(Ben-Shachar, 2023)
Machine Learning	caret, randomForest, xgboost	Predictive modeling with machine learning techniques	Random Forest, XGBoost	Handling large datasets and computational complexity	Validate model performance with cross-validation	(Nandipati & Boddala, 2024)
PK/PD Modeling	nlme, PKNCA, mrgsolve	Modeling drug pharmacokinetics and dynamics	PK/PD Modeling	Assumption of compartment models	Ensure correct parameterization for different compartments	(Sood & Anita, 2024)
Bioequivalence Testing	TOSTER	Testing bioequivalence between drug formulations	Two One-Sided T-Tests	Selecting appropriate thresholds for bioequivalence	Set bioequivalence acceptance range (80-125%)	(Peck et al., 2022)

7. Data Interpretation and Reporting

7.1. Interpreting Outputs of Statistical Models

Understanding statistical models is a prerequisite to transform numbers into useful knowledge in drug-related studies. The main considerations are confidence intervals, p-values, and effect sizes. Confidence intervals represent a distribution range for a true parameter value with large samples values nearing the true value with limited possibilities of error margins or confidence intervals being accurate estimates of the uncertainity. P-values are regarded as probabilities that the conclusion may likely just be by chance, adjusted after keeping into consideration various effects of p-values typically .05 or higher not considered, statistically speaking for most drug studies. Effect sizes assess whether differences or relationships between groups vary in magnitude to determine if the findings are practically significant. Each of these three metrics independently inform researchers about the strength and importance of their results, and together they are more useful to researchers for accurately reporting statistics (Mweshi & Sakyi, 2020).

7.2. Graphical Representation of Model Results

To visualize model outcomes will improve interpretation and communication of statistical results in pharmacology. Forest plots are commonly used to present results from meta-analyses or regression models (for example, with random-effects models), with estimates and confidence intervals on the same display for effect sizes across different studies or covariates. ROC curves are useful for evaluating classification models or comparing the performance of classification models, as they clearly show the trade-off between sensitivity and specificity at varying cut-off thresholds. Prediction plots, like those comparing predicted to observed outcomes, are important for assessing model predictive accuracy and how well the model fits the data. Graphical components like those discussed, make complex statistical output more intuitive, accessible, and easier to understand (Chan et al., 2024).

7.3. Best Practices for Reproducibility

Reproducibility is an important consideration in pharmacological research to give confidence that the results can be consistently reproduced. R Markdown allows researchers to combine code, analysis, and results within the same document, allowing greater transparency, as well as making it easier to update the document in the future. It also allows researchers to create dynamic and reproducible reports for sharing with colleagues and collaborators, and it provides a simple way to update those reports. Version control systems (e.g. Git) track changes made to code, data, and analyses, facilitating sharing of code and analysis between people and for collaboration, collaboration between people on the project, and even the possibility of reverting back to previous versions of documents because of the version history tracked. These practices will protect the integrity of the research as well as improve team collaboration and workflow (Schaduangrat et al., 2020).

8. Case Studies and Examples

8.1. Dose-response Curve Analysis

Dose-response curve analysis is a method used in pharmacology to research the relationship between drug dosage and pharmacological outcome. In research studies when Scientists conduct a research study to look at a new drug's pharmacology, they will assess the effectiveness of that drug in studies where they use a range of drug concentrations and report an effect. Using that information, they will assess the dose-response curve and use it to determine important parameters such as the Emax, the ED50 (that is the dose producing 50% of Emax), and the slope of the dose-response curve. Using statistical modeling approaches (such as the drc package in R) researchers will quantitatively assess the potency of the drug, e.g. by determining appropriate dosing regimens to be implemented during clinical studies. Therefore, dose-response curve analysis is a critical tool for pharmacologists to identify plausible evidence based safe and efficacious dose levels during drug development (Moffett et al., 2022).

8.2. Comparative Efficacy of drug Formulations

Comparative efficacy studies are significant in comparing alternative drug formulations to measure which formulation produces the greatest therapeutic effect. For example, establishing the pharmacokinetics (Cmax and AUC) of a new generic formulation in comparison to an original branded drug by using a case study design. The appropriate statistical analysis would be a two-sample t-test or ANOVA testing for significant differences in efficacy between formulations. Such analyses determine if the new formulation is bioequivalent and provides similar therapeutic outcomes as the original branded drug to ultimately assist with regulatory approval and clinician decision making (van der Koog et al., 2022).

8.3. Real-world Pharmacovigilance Dataset Analysis

Real-world data analysis is important for monitoring adverse reactions (safety) and efficacy after drugs have gone to market. Researchers used various statistical techniques to analyze reports of adverse reactions, ultimately used to detect possible risks or emerging signals related to drug exposure. For example, survival analysis and logistic regression allows researchers to estimate the probability of side effects according to the given variables for rate or amount of dose, patient information as demographics, and pre-existing health conditions, etc. Data visualization techniques such bar charts, histograms, and time series plots will often provide useful information for observing patterns or sudden increases in reports of side effects. This type of analysis represents the third mechanism of pecunious drug surveillance and can expedite protecting public health (Dang, 2023).

9. Limitations and Challenges

9.1. Statistical Assumptions and Model fit

Within pharmacological data analysis, the use of statistical assumptions and model fitting can become difficult. Many commonly used

statistical methods like linear regression and ANOVA rely on several assumptions like normality, independence, and homogeneity of variance, so violations of these assumptions can produce biased or misleading estimates. Therefore, it is critical to appropriately fit the model to the data for accurate interpretations. While accurately fitting a model to the data can sometimes be easier said than done, especially when dealing with nonlinear associations or heterogeneous data, model diagnostics can help evaluate model fit yet may still face issues such as skewed data, outliers, or missing values. Addressing the concerns as noted is important so that valid conclusions can be made (McComb et al., 2022).

9.2. Data Quality and Sampling Issues

Issues related to data quality and sampling pose challenges to pharmacological research. Data sources that are poor quality such as estimates, nonexistent information, or biased data can potentially jeopardize the credibility of the statistical analysis portion of the research. Other problems may involve sample sizes being small enough that they do not provide sufficient statistical power for making inferences on a meaningful effect or difference. In addition, sampling in a non-representative way a sample having little to no demographic considerations -- can introduce bias and again limit the ability to generalize as the findings can merely apply to those sampled. Use random lists of representative individuals to minimize bias and reporting with appropriate sampling methods will improve your final results; representation should include their demographics. Rigorous cleaning and validation of the data are necessary steps to take and may be required prior to the use of high quality data (Elliott, 2020).

9.3. Interpretation Pitfalls in Pharmacological Context

Errors in interpretation in clinical drug research can lead to incorrect conclusions and ill-informed decisions. One easy error is to make too much of results that come from small sample sizes, resulting in conclusions based on too little power to be correct. There is also a familiar problem with mixing correlation and causation; just because two variables have a relationship does not mean that one causes the other. Likewise, the presence of errors can lead to misunderstanding if important variables are neglected or confounding biases are not corrected. In pharmacology, asking questions about the dose-response relationship or population variability could also imperil error. Accurate data review, usage of appropriate models, and proper interpretation will go a long way toward avoiding these errors (Hansson, 2020).

Conclusion

Statistical concepts are important for establishing validity and reliability in pharmacological research. Important aspects to consider are understanding the assumptions of the statistical models used, addressing any quality issues with the data collected, and selecting the appropriate techniques for hypothesis testing (for example, t-tests, ANOVAs), correlation, and regression analysis (general linear modelling) or advanced forms of regression analysis. R Studio is a wonderful platform for dealing with pharmacological data and performing various forms of analyses, especially considering the superior statistical features it has (we have used it in our studies) and the availability of particular packages (e.g. drc, lme4, caret) designed for handling complex data. Although it can be advanced, R Studio has a lot of advantages, including flexibility, reproducibility, and connectivity to visualization packages you can use to explore your data in many reasonable ways. Overall, I see a large potential role for AI and real-world data in pharmacological thinking. In the future, machine learning can identify patterns across large and often complex datasets, inform future research directions, and inform practical solutions to drug efficacy, safety, and personalized approaches to treatments. Similarly, real-world evidence, such as electronic health records and patient-reported and patient-related outcomes, can provide richer qualitative and quantitative data related to drug effects in environments that are not so tightly controlled as clinical practice and pre-clinical evaluation. As technology continues to advance your thinking in pharmacology will become more accurate, efficient, and personalized when drug development is completed and practiced.

References

Alabi, O., & Bukola, T. (2023). Introduction to Descriptive statistics. In *Recent Advances in Biostatistics*. IntechOpen.

Altaf, S., & Iqbal, T. (2023). Bee Venom Used for the Treatment of Rheumatoid Arthritis. *Biomedical Journal of Scientific & Technical Research*, *53*(2), 44503-44507.

Altaf, S., Iqbal, T., Majeed, W., Farooq, M. A., Naseer, D., Saleem, M., Babar, S. U. R., & Ikram, M. (2023). Plasma membrane camouflaged nanoparticles: an emerging antibacterial approach. *One Health Triad, Unique Scientific Publishers, Faisalabad, Pakistan*, *2*, 193-200.

Altaf, S., Iqbal, T., Salma, U., Sajid, M., Basit, I., Sabir, M. Z., Riaz, K., Rasheed, R., Umair, M., & Talha, R. (2024). *Gold nanoparticles for the detection of organophosphate. Agrobiological Records 16: 11-18*.

Bao, Y., Zhou, H., Fu, Y., Wang, C., & Huang, Q. (2024). Zhumian Granules improves PCPA-induced insomnia by regulating the expression level of neurotransmitters and reducing neuronal apoptosis. *Journal of Ethnopharmacology*, *327*, 118048.

Bayman, E. O., & Dexter, F. (2021). Multicollinearity in logistic regression models. In *Anesthesia & Analgesia* (Vol. 133, Issue 2, pp. 362-365). LWW.

Ben-Shachar, M. S. (2023). *Order constraints in Bayes models (with brms)*.

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920960351.

Bulanov, N. M., Suvorov, A. Y., Blyuss, O. B., Munblit, D. B., Butnaru, D. V, Nadinskaia, M. Y., & Zaikin, A. A. (2021). Basic principles of descriptive statistics in medical research. *Сеченовский Вестник*, *12*(3), 4-16.

Chan, J. R., Allen, R., Boras, B., Cabal, A., Damian, V., Gibbons, F. D., Gulati, A., Hosseini, I., Kearns, J. D., & Saito, R. (2024). Current practices for QSP model assessment: an IQ consortium survey. *Journal of Pharmacokinetics and Pharmacodynamics*, *51*(5), 543-555.

Charoo, N. A. (2020). Converging generic drug product development: bioequivalence design and reference product selection. *Clinical Pharmacokinetics*, *59*(11), 1335-1355.

Coleman, M. D. (2020). *Human drug metabolism*. John Wiley & Sons.

Dagenais, S., Russo, L., Madsen, A., Webster, J., & Becnel, L. (2022). Use of real-world evidence to drive drug development strategy and inform clinical trial design. *Clinical Pharmacology & Therapeutics*, *111*(1), 77–89.

Dang, A. (2023). Real-world evidence: a primer. *Pharmaceutical Medicine*, *37*(1), 25–36.

Elkomy, M. H., Hendawy, O. M., Zaki, R. M., Tulbah, A. S., Aldosari, B. N., Ali, A. A., & Eid, H. M. (2024). Intranasal trimethyl chitosan-coated emulsomes containing tizanidine as brain-targeted therapy in spasticity: formulation, optimization, and pharmacokinetic assessment. *Drug Delivery and Translational Research*, 1–16.

Elliott, S. (2020). Impact of inadequate methods and data analysis on reproducibility. *Journal of Pharmaceutical Sciences*, *109*(2), 1211–1219.

Fu, J., Zhang, Y., Liu, J., Lian, X., Tang, J., & Zhu, F. (2021). Pharmacometabonomics: data processing and statistical analysis. *Briefings in Bioinformatics*, *22*(5), bbab138.

Gerard, E. (2021). *Bayesian toxicity modeling for dose regimen assessment in early phase oncology trials incorporating pharmacokinetics and pharmacodynamics*. Université Paris Cité.

Hansson, S. O. (2020). Values in pharmacology. *Uncertainty in Pharmacology: Epistemology, Methods, and Decisions*, 375–396.

Hedaya, M. A. (2023). *Basic pharmacokinetics*. Routledge.

Iqbal, T., Ahmad, A., Naveed, M. T., Ali, A., & Ahmad, M. (2023). Potential Role of Zoonoses in Bioterrorism. *Zoonosis, Unique Scientific Publishers, Faisalabad, Pakistan*, *1*, 499–512.

Iqbal, T., Altaf, S., Basit, I., Naeem, M. A., Akram, Q., Saeed, M. R., Hyder, S., & Salma, U. (2024). Hesperetin: A Potent Phytochemical Constituent for the Treatment of Rheumatoid Arthritis: Hesperetin for the Treatment of Rheumatoid Arthritis. *Pakistan BioMedical Journal*, 2–10.

Iqbal, T., Altaf, S., Salma, U., Fatima, M., Khan, M. N., Farooq, S., Abrar, M., Tasleem, M., & Afzal, A. (2024). *Cell membrane coated polymeric nanocarriers: a novel drug delivery approach for the targeted therapy of rheumatoid arthritis. Agrobiological Records 15: 91-102*.

Jarantow, S. W., Pisors, E. D., & Chiu, M. L. (2023). Introduction to the use of linear and nonlinear regression analysis in quantitative biological assays. *Current Protocols*, *3*(6), e801.

Jordan, D. G., Winer, E. S., & Salem, T. (2020). The current status of temporal network analysis for clinical science: Considerations as the paradigm shifts? *Journal of Clinical Psychology*, *76*(9), 1591–1612.

Kabacoff, R. (2024). *Modern data visualization with R*. CRC Press.

Kim, M. S., An, M. H., Kim, W. J., & Hwang, T.-H. (2020). Comparative efficacy and safety of pharmacological interventions for the treatment of COVID-19: A systematic review and network meta-analysis. *PLoS Medicine*, *17*(12), e1003501.

Kuhn, M., & Silge, J. (2022). *Tidy modeling with R: A framework for modeling in the tidyverse*. " O'Reilly Media, Inc."

Kumar-M, P., & Mishra, A. (2024). Inferential Statistics for the Hypothesis Testing of Non-parametric Data. In *R for Basic Biostatistics in Medical Research* (pp. 157–197). Springer.

Mahoney, M. (2021). *Lampreys adjust, mammals non-plussed, birds robust: how ecological and environmental features influence genetic diversity*.

McComb, M., Bies, R., & Ramanathan, M. (2022). Machine learning in pharmacometrics: Opportunities and challenges. *British Journal of Clinical Pharmacology*, *88*(4), 1482–1499.

Michel, M. C., Murphy, T. J., & Motulsky, H. J. (2020). New author guidelines for displaying data and reporting data analysis and statistical methods in experimental biology. *Molecular Pharmacology*, *97*(1), 49–60.

Miranda, M., Veloso, C., Brown, M., Pais, A. A. C. C., Cardoso, C., & Vitorino, C. (2022). Topical bioequivalence: Experimental and regulatory considerations following formulation complexity. *International Journal of Pharmaceutics*, *620*, 121705.

Moffett, D. B., Mumtaz, M. M., Sullivan Jr, D. W., & Whittaker, M. H. (2022). General considerations of dose-effect and dose-response relationships. In *Handbook on the Toxicology of Metals* (pp. 299–317). Elsevier.

Monkman, M. H. (2024). *The Data Preparation Journey: Finding Your Way with R*. CRC Press.

Mweshi, G. K., & Sakyi, K. (2020). Application of sampling methods for the research design. *Archives of Business Review–Vol*, *8*(11), 180–193.

Myint, L., Hadavand, A., Jager, L., & Leek, J. (2020). Comparison of beginning R students' perceptions of peer-made plots created in two plotting systems: a randomized experiment. *Journal of Statistics Education*, *28*(1), 98–108.

Nandipati, V. S. S., & Boddala, L. V. (2024). *Credit Card Approval Prediction: A comparative analysis between Logistic Regression, KNN, Decision Trees, Random Forest, XGBoost*.

Obaido, G., Mienye, I. D., Egbelowo, O. F., Emmanuel, I. D., Ogunleye, A., Ogbuokiri, B., Mienye, P., & Aruleba, K. (2024). Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, and prospects. *Machine Learning with Applications*, *17*, 100576.

Ohlsson, H., & Kendler, K. S. (2020). Applying causal inference methods in psychiatric epidemiology: A review. *JAMA Psychiatry*, *77*(6), 637–644.

Okoye, K., & Hosseini, S. (2024). *R Programming: Statistical Data Analysis in Research*. Springer Nature.

Olsen, A. A., McLaughlin, J. E., & Harpe, S. E. (2020). Using multiple linear regression in pharmacy education scholarship. *Currents in Pharmacy Teaching and Learning*, *12*(10), 1258–1268.

Ota, R., & Yamashita, F. (2022). Application of machine learning techniques to the analysis and prediction of drug pharmacokinetics. *Journal of Controlled Release*, *352*, 961–969.

Peck, C., Campbell, G., Yoo, I., Feng, K., Hu, M., & Zhao, L. (2022). Comparing a Bayesian Approach (BEST) with the Two One-Sided t-Tests (TOSTs) for Bioequivalence Studies. *The AAPS Journal*, *24*(5), 97.

Ruberg, S. J., Beckers, F., Hemmings, R., Honig, P., Irony, T., LaVange, L., Lieberman, G., Mayne, J., & Moscicki, R. (2023). Application of Bayesian approaches in drug development: starting a virtuous cycle. *Nature Reviews Drug Discovery*, *22*(3), 235–250.

Schaduangrat, N., Lampa, S., Simeon, S., Gleeson, M. P., Spjuth, O., & Nantasenamat, C. (2020). Towards reproducible computational drug discovery. *Journal of Cheminformatics*, *12*, 1–30.

Schwartz, S. M., Wildenhaus, K., Bucher, A., & Byrd, B. (2020). Digital twins and the emerging science of self: implications for digital health experience design and "small" data. *Frontiers in Computer Science*, *2*, 31.

Shatz, I. (2024). Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics. *Behavior Research Methods*, *56*(2), 826–845.

Sood, R., & Anita, A. (2024). Pharmacokinetic and Pharmacodynamic Modeling (PK/PD) in Pharmaceutical Research: Current Research and Advances. *Software and Programming Tools in Pharmaceutical Research*, 153–169.

van der Koog, L., Gandek, T. B., & Nagelkerke, A. (2022). Comparative efficiency studies are essential for evaluating various drug formulations and destinations that provide the best therapeutic outcome. *Advanced Healthcare Materials*, *11*(5), 2100639.

Zhang, Y., & Thorburn, P. J. (2022). Handling missing data in near real-time environmental monitoring: A system and a review of selected methods. *Future Generation Computer Systems*, *128*, 63–72.

Zhao, X., Iqbal, S., Valdes, I. L., Dresser, M., & Girish, S. (2022). Integrating real-world data to accelerate and guide drug development: A clinical pharmacology perspective. *Clinical and Translational Science*, *15*(10), 2293–2302.

Zou, H., Banerjee, P., Leung, S. S. Y., & Yan, X. (2020). Application of pharmacokinetic-pharmacodynamic modeling in drug delivery: development and challenges. *Frontiers in Pharmacology*, *11*, 997