World Journal of Biology Pharmacy and Health Sciences
eISSN: 2582-5542
Cross Ref DOI: 10.30574/wjbphs
Journal homepage: https://wjbphs.com/

(REVIEW ARTICLE)

Statistical challenges and solutions in multidisciplinary clinical research: Bridging the gap between

Opeyemi Olaoluawa Ojo [1,*] and Blessing Kiobel [2]

[1] *Tritek Business Consulting, London United Kingdom.*
[2] *College of Nursing, Xavier University, Ohio, USA.*

World Journal of Biology Pharmacy and Health Sciences, 2024, 19(03), 246–258

Publication history: Received on 01 August 2024; revised on 08 September 2024; accepted on 11 September 2024

Article DOI: https://doi.org/10.30574/wjbphs.2024.19.3.0628

Abstract

This paper delves into the intricate challenges and innovative solutions in applying statistical methodologies within clinical research, aiming to bridge the gap between biostatistics and medicine. The study meticulously examines fundamental biostatistical concepts, addressing the complexities of modern clinical trials and observational studies. Through a comprehensive review of advanced regression models, causal inference techniques, and machine learning algorithms, the paper illuminates the evolving landscape of biostatistics in handling high-dimensional data and confounding variables.

The methods employed in this study involve an extensive analysis of current literature, case studies, and practical applications that demonstrate the utility of these advanced methodologies. Key findings reveal that traditional statistical approaches often fall short in capturing the complexities of clinical data, necessitating the adoption of more sophisticated techniques. The integration of non-linear regression models, robust causal inference methods, and machine learning has significantly enhanced the accuracy and reliability of research outcomes, offering deeper insights into patient outcomes and treatment efficacy.

Conclusions drawn from this study underscore the critical need for a paradigm shift in clinical research, moving beyond the rigid reliance on p-values towards a more holistic approach that emphasizes effect sizes, confidence intervals, and practical significance. The paper recommends continued innovation in statistical methodologies, particularly the integration of big data analytics and machine learning, to address the growing complexities of biomedical data. Furthermore, it advocates for interdisciplinary collaboration and ethical considerations in the application of these advanced techniques to ensure that biostatistics continues to contribute meaningfully to the advancement of medical science.

Keywords: Biostatistics; Causal Inference; Machine Learning; Clinical Research; Advanced Regression Models; Big Data Analytics

1. Introduction

The integration of biostatistics into clinical research has become increasingly pivotal as the complexity and scope of clinical studies expand across multiple disciplines. Traditionally, clinical research and biostatistics were seen as parallel but distinct domains; however, the need for a more cohesive and interdisciplinary approach has emerged as essential for addressing the intricate challenges posed by modern medicine. Clinical research now encompasses a broader spectrum of scientific inquiry, requiring not only statistical expertise but also an understanding of the clinical context in which data is generated and interpreted (Freedman, Lowe & Macaskill, 1984). This interdisciplinary approach is

* Corresponding author: Opeyemi Olaoluawa Ojo

further exemplified in environmental studies, where the impact of factors like ultra-violet light radiation on microbial growth kinetics and biodegradation performance has been rigorously studied to understand the broader implications for biostatistics in environmental health research (Eregie et al., 2024).

Biostatistics plays a critical role in the design, execution, and analysis of clinical trials, providing the mathematical framework necessary to ensure the validity and reliability of study outcomes. The evolution of clinical trials, from simple comparative studies to complex, adaptive trials, has necessitated the development of advanced statistical methods that can accommodate the multifaceted nature of clinical data (Pocock & Simon, 1975). This complexity is further amplified by the fragility of interconnected technological systems, as demonstrated by recent events (Ogundipe & Aweto, 2024)

These The challenges in this field are diverse and include determining the appropriate sample size, employing effective randomization techniques, managing missing data, and ensuring that the interpretation of results is relevant and useful to both clinicians and policymakers (Signorini et al., 1993; NAIIS, 2018).

One of the central challenges in bridging the gap between biostatistics and clinical practice is the issue of communication and collaboration. Clinicians and statisticians often operate within different paradigms, with clinicians focusing on the practical implications of research findings and statisticians emphasizing methodological rigor (Stensrud et al., 2022). This disconnect can lead to misunderstandings and misapplications of statistical methods in clinical settings, potentially compromising patient care. It is therefore crucial to foster a collaborative environment where both statisticians and clinicians can contribute their expertise to the research process (Joseph & Uzondu, 2024).

Furthermore, the introduction of new technologies such as artificial intelligence (AI) and machine learning (ML) into clinical research has added another layer of complexity to the statistical challenges faced by researchers. These technologies offer powerful tools for data analysis, enabling the identification of patterns and relationships that might not be apparent through traditional statistical methods. However, the integration of AI and ML into clinical research also raises significant challenges, including the need for new statistical models that can handle large, complex datasets and the ethical considerations associated with the use of automated decision-making tools in healthcare (Joseph & Uzondu, 2024).

The use of AI and ML in clinical research has also highlighted the importance of interdisciplinary education and training. As the boundaries between biostatistics, data science, and clinical medicine continue to blur, there is a growing need for educational programs that equip researchers with the skills necessary to navigate this increasingly complex landscape. The development of curricula that integrate these disciplines is crucial for preparing the next generation of researchers to tackle the statistical challenges of modern clinical research (Joseph & Uzondu, 2024).

In addition to these educational initiatives, there is a pressing need for more robust statistical methods that can address the unique challenges posed by multidisciplinary clinical research. For example, the use of survival analysis in clinical trials has become increasingly common, yet traditional methods may not be adequate for dealing with the complexities of modern clinical data. Advanced methods such as counting processes and dynamic randomization techniques are being developed to better account for the intricacies of clinical trial data and improve the accuracy and reliability of study outcomes (Fleming & Harrington, 2013; Pocock & Simon, 1975).

Moreover, the issue of bias in clinical research remains a significant concern. Bias can arise at various stages of a clinical study, from the selection of participants to the interpretation of results, and can severely compromise the validity of the research. Addressing bias requires a combination of rigorous study design, careful data management, and sophisticated statistical techniques (Pannucci & Wilkins, 2010; Martínez-Mesa et al., 2016). For instance, Hayden et al. (2013) discuss the importance of assessing bias in prognostic studies, emphasizing the need for statistical methods that can identify and adjust for potential sources of bias in clinical research.

Another critical aspect of modern clinical research is the move towards personalized medicine, which aims to tailor treatments to individual patients based on their genetic, environmental, and lifestyle factors. This approach requires the development of new statistical methods that can handle the vast amounts of data generated by personalized medicine studies and translate these data into actionable clinical insights. The work of Harrell (2012) on regression modeling strategies provides a foundation for developing these methods, offering tools for analyzing complex, high-dimensional data and identifying the most relevant predictors of clinical outcomes.

The debate over the use of statistical significance in clinical research also continues to be a contentious issue. While traditional statistical methods often rely on p-values to determine the significance of study results, there is growing recognition that this approach has significant limitations. McShane et al. (2019) argue for abandoning statistical

significance in favor of more nuanced methods that take into account the broader context of the research and the practical implications of the findings. This shift towards a more holistic approach to statistical analysis is essential for ensuring that clinical research produces results that are both scientifically valid and clinically relevant.

This study aims to explore the challenges and propose solutions that enhance the quality and impact of clinical research through the integration of biostatistics. The objective is to investigate how new statistical methods, combined with collaborative and interdisciplinary approaches, can address the complex challenges in modern clinical studies. By focusing on the intersection of biostatistics and clinical practice, this research will contribute to the development of robust statistical tools and frameworks that ensure the reliability and applicability of clinical research outcomes. The scope of this study includes an examination of advanced statistical techniques, the role of technology in clinical research, and the importance of interdisciplinary education and collaboration in bridging the gap between biostatistics and medicine.

2. Fundamental Concepts in Biostatistics

Biostatistics is a critical field that underpins the scientific rigor of clinical research, providing the mathematical tools necessary for the design, analysis, and interpretation of biomedical studies. As clinical research evolves to encompass more complex and interdisciplinary investigations, a solid understanding of fundamental biostatistical concepts becomes indispensable for both researchers and clinicians. These concepts ensure that clinical trials and observational studies yield valid, reliable, and interpretable results, which are essential for advancing medical science and improving patient outcomes (Harrell, 2012). Furthermore, the application of biostatistical methods extends beyond clinical trials to environmental research, as seen in studies comparing the biodegradative efficiency of wildtype versus mutagenised Scenedesmus vacuolatus strains on spent coolant waste. These studies highlight the role of biostatistics in analyzing dehydrogenase activity and petroleum degradation, demonstrating its broad applicability in various scientific domains (Eregie and Jamal-Ally, 2023).

One of the core concepts in biostatistics is the design of experiments, particularly the use of randomization to reduce bias. Randomization is a key element in clinical trials, as it ensures that treatment groups are comparable and that the results are not influenced by confounding factors. Pocock and Simon (1975) introduced sequential treatment assignment with balancing for prognostic factors, a method that aims to maintain balance between groups on important covariates while randomizing participants. This approach minimizes selection bias and enhances the validity of the trial results.

Another fundamental aspect of biostatistics is the analysis of survival data, which is particularly relevant in clinical trials involving time-to-event outcomes. Fleming and Harrington (2013) discuss counting processes and survival analysis, which provide the statistical framework for analyzing data where the outcome is the time until an event occurs, such as death or disease progression. Survival analysis techniques, such as the Kaplan-Meier estimator and the Cox proportional hazards model, are widely used to estimate survival probabilities and to assess the effect of covariates on survival times (Clark et al., 2003; Schober et al., 2018).

The issue of bias is a recurring concern in clinical research, as it can severely compromise the validity of study findings. Hayden et al. (2013) emphasize the importance of assessing and addressing bias in studies of prognostic factors, noting that bias can arise at various stages of a study, from participant selection to data analysis. Biostatisticians utilize several methods to reduce bias, including stratified randomization and multivariable modeling to control for confounding factors. Signorini et al. (1993) discuss the comparison between stratified randomization and dynamic balancing, emphasizing the strengths of each approach in minimizing bias and enhancing the reliability of clinical trial outcomes.

The interpretation of statistical significance is a topic of ongoing debate in the field of biostatistics. McShane et al. (2019) argue for abandoning the traditional reliance on p-values as the sole criterion for statistical significance, advocating instead for a more nuanced approach that considers the broader context of the research findings. This shift towards a more comprehensive interpretation of statistical results is crucial for ensuring that clinical research provides meaningful insights that can inform clinical practice.

Biostatistics also plays a central role in the design and analysis of observational studies, which are often used to investigate the natural history of diseases and the effects of exposures on health outcomes. Schwartz and Lellouch (1967) discuss the distinction between explanatory and pragmatic attitudes in therapeutic trials, highlighting the different objectives of these study designs. Explanatory trials are designed to test hypotheses under ideal conditions, while pragmatic trials aim to evaluate the effectiveness of interventions in real-world settings. Understanding these

distinctions is important for designing studies that are fit for their intended purpose and for interpreting their results in the appropriate context (Patsopoulos, 2011; Borgerson, 2013).

The development of regression models is another cornerstone of biostatistics, providing the tools necessary for analyzing the relationships between multiple variables in clinical research. Harrell (2012) offers comprehensive guidance on regression modeling strategies, covering applications in linear models, logistic regression, ordinal regression, and survival analysis. These models are essential for understanding the complex interplay between different factors in clinical studies and for making accurate predictions about patient outcomes.

Biostatistics also addresses the challenges of analyzing data from complex clinical trials, where multiple endpoints and interim analyses are common. Bland and Altman (1986) provide statistical methods for assessing agreement between different measurement methods, which are crucial for ensuring the reliability of clinical trial results. These methods help to identify and quantify the degree of agreement between different assessments, reducing the likelihood of measurement error and improving the precision of study findings.

Finally, the professional development of researchers in biostatistics is essential for ensuring that clinical research remains at the forefront of scientific innovation. Joseph and Uzondu (2024) discuss the importance of continuous learning and professional development for STEM educators, emphasizing the need for ongoing training in advanced statistical methods. This commitment to education and professional growth is vital for maintaining the high standards of biostatistical practice and for fostering a culture of excellence in clinical research.

2.1. Challenges in Applying Statistical Methods in Clinical Research 4.1

The application of statistical methods in clinical research presents a myriad of challenges that can significantly impact the validity and reliability of study outcomes. As clinical trials and observational studies become more complex, the statistical tools and methodologies employed must evolve to address the intricacies inherent in biomedical data. These challenges are multifaceted, encompassing issues related to study design, data collection, analysis, and interpretation. Each stage of clinical research is fraught with potential pitfalls that can lead to bias, errors and misinterpretation if not properly managed (Pocock & Simon, 1975; Layode et al. 2024a).

One of the fundamental challenges in applying statistical methods in clinical research is the design of experiments, particularly the process of randomization. Randomization is a critical component of clinical trials, as it minimizes selection bias and ensures that the treatment groups are comparable (NAIIS, 2018). However, achieving true randomization in practice can be difficult, especially in studies with small sample sizes or complex stratification requirements. Pocock and Simon (1975) addressed these challenges by developing sequential treatment assignment methods that balance prognostic factors, thereby enhancing the reliability of the randomization process. Despite these advances, achieving balance and minimizing bias remains a significant challenge in many clinical trials.

Another key challenge is the assessment of agreement between different methods of clinical measurement. In many studies, researchers must rely on multiple instruments or techniques to measure outcomes, which can lead to discrepancies in the data. Bland and Altman (1996) developed statistical methods for assessing agreement between two methods of clinical measurement, providing a framework for quantifying the consistency between different measurement techniques. However, the application of these methods requires careful consideration of the underlying assumptions and potential sources of measurement error, which can complicate the interpretation of results.

Causal inference is another area where statistical methods face significant challenges in clinical research. In randomized controlled trials (RCTs), the randomization process helps to establish causality by ensuring that the treatment and control groups are comparable. However, in observational studies where randomization is not possible, establishing causality becomes more complex. Stensrud et al. (2022) discuss advanced methods for causal inference, such as propensity score matching and instrumental variable analysis, which aim to approximate the conditions of an RCT in observational data. These methods are powerful tools for drawing valid conclusions about cause-and-effect relationships, but they are also prone to biases and require careful application to avoid misleading results.

The interpretation of statistical significance is another contentious issue in clinical research. Traditionally, p-values have been used as the primary criterion for determining whether a result is statistically significant. However, this approach has been increasingly criticized for its limitations, particularly its tendency to oversimplify complex data and overlook the broader context of the research findings. McShane et al. (2019) argue for abandoning the rigid use of p-values in favor of a more nuanced approach that considers the practical significance and clinical relevance of the results.

This shift requires researchers to adopt a more holistic perspective on statistical analysis, which can be challenging in a field that has long relied on p-values as a benchmark for success.

Bias is another pervasive challenge in clinical research, affecting all stages of the research process from study design to data analysis. Bias can arise from a variety of sources, including selection bias, measurement bias, and confounding. Hayden et al. (2013) emphasize the importance of assessing and mitigating bias in studies of prognostic factors, noting that even small biases can significantly distort study results. The use of advanced statistical techniques, such as stratified randomization and multivariable adjustment, can help to reduce bias, but these methods are not foolproof and require careful implementation to be effective.

The complexity of clinical data is another significant challenge in applying statistical methods. Clinical studies often involve large datasets with multiple variables, making it difficult to identify the most relevant predictors and interactions (Jahun et al, 2021). Harrell (2012) discusses regression modeling strategies that can be used to address these challenges, offering tools for analyzing complex, high-dimensional data. However, these models are often highly sensitive to assumptions about the data, and incorrect model specification can lead to biased estimates and incorrect conclusions. The challenge of balancing explanatory and pragmatic attitudes in clinical trials also complicates the application of statistical methods. Schwartz and Lellouch (1967) distinguish between explanatory trials, which aim to test hypotheses under controlled conditions and pragmatic trials, which evaluate interventions in real-world settings.

The statistical methods used in these different types of trials must be tailored to the specific objectives of the study. For example, while explanatory trials may rely on strict randomization and controlled conditions, pragmatic trials may require more flexible methods that account for variability in clinical practice. This distinction highlights the importance of aligning statistical methods with the goals of the study, which is not always straightforward (Casey et al., 2022; Tashkin, 2020; Seyi- Lande).

The use of survival analysis in clinical research introduces additional challenges, particularly in studies where the outcome of interest is the time until an event occurs. Fleming and Harrington (2013) discuss counting processes and survival analysis, which are essential tools for analyzing time-to-event data. However, these methods are complex and require careful handling of censoring and time-dependent covariates. Moreover, the assumptions underlying survival analysis methods, such as the proportional hazards assumption in the Cox model, may not always hold in practice, leading to biased results if not properly addressed (Schober et al., 2018; Sloan & Dueck, 2004).

Stratified randomization and dynamic balancing are techniques used to enhance the validity of clinical trials by ensuring that important covariates are balanced across treatment groups. Signorini et al. (1903) compare these methods, noting that while both can effectively reduce bias, they also introduce new challenges. For example, dynamic balancing methods may increase the risk of selection bias if not properly implemented and stratified randomization requires careful planning to ensure that all relevant covariates are accounted for. These challenges highlight the trade-offs involved in applying advanced statistical methods in clinical research (Signorini, 1993; Kang et al., 2008).

Lastly, the integration of digital technologies and data-driven approaches in clinical research introduces new challenges in the application of statistical methods. Joseph and Uzondu (2024) discuss the role of digital tools in bridging the digital divide in STEM education, noting that while these technologies offer powerful new capabilities for data analysis, they also require researchers to develop new skills and knowledge. The same applies to clinical research, where the use of big data and machine learning techniques demands a deep understanding of both the technology and the underlying statistical principles. As noted by Schober et al. (2018), the complexity of these methods can lead to challenges in ensuring the validity and reliability of research findings. Additionally, Perera et al. (2020) highlight the necessity of addressing these challenges to fully realize the potential of these advanced methodologies in clinical research.

2.2. Innovative Statistical Solutions and Methodologies

The rapid evolution of clinical research necessitates the development and application of innovative statistical solutions and methodologies to address the increasingly complex challenges presented by biomedical data. Traditional statistical approaches, while foundational, often fall short in handling the nuances of modern clinical trials and observational studies (Harrell, 2001: Fahr, P., Buchanan & Wordsworth, 2019). As a result, statisticians and researchers have devised advanced techniques that not only improve the accuracy and validity of clinical research findings but also enhance their interpretability and relevance in clinical practice (Harrell, 2012; Layode et al 2024b).

One of the key innovations in statistical methodologies is the development of advanced regression techniques that extend beyond the traditional linear models. Harrell (2012) provides a comprehensive framework for regression

modeling strategies, which includes logistic regression, ordinal regression, and survival analysis. These methods are particularly useful in clinical research where outcomes are often categorical or time-dependent. By incorporating advanced regression techniques, researchers can more accurately model the relationships between multiple variables and account for the complexity of clinical data. (Serdar et al., 2021; Lee et al., 2019).

Causal inference remains a cornerstone of clinical research, particularly in observational studies where randomization is not feasible. Stensrud et al. (2022) discuss the application of advanced causal inference methods, such as marginal structural models and inverse probability weighting, which allow researchers to estimate causal effects in the presence of time-varying confounders. These methodologies are critical for drawing valid conclusions about cause-and-effect relationships in clinical research, thereby bridging the gap between observational studies and randomized controlled trials (RCTs).

The shift away from strict reliance on p-values as the sole measure of statistical significance has led to the adoption of more nuanced approaches to statistical analysis. McShane et al. (2019) argue for abandoning the traditional use of p-values in favor of approaches that consider the practical significance and broader context of research findings. This paradigm shift encourages researchers to focus on the effect sizes, confidence intervals, and the overall impact of their findings rather than merely achieving statistical significance. Such a shift is essential for enhancing the clinical relevance of research outcomes. (Davis et al., 2021; Armijo-Olivo, 2018).

Missing data is a common challenge in clinical research, and innovative solutions have been developed to address this issue. Multiple imputation, as proposed by Rubin (2004), has become a standard method for handling missing data in clinical trials and observational studies. This approach involves generating several datasets by filling in missing values using observed data, followed by separate analyses of each dataset. The results are then combined to produce accurate statistical conclusions. Enders (2017) elaborates on advanced imputation techniques that enhance the management of missing data in complex datasets. These strategies are essential for preserving the reliability of clinical research outcomes despite the challenges posed by incomplete data.

Another significant advancement in statistical methodologies is the development of regularization techniques for variable selection and model building. Tibshirani (1996) introduced the Lasso (Least Absolute Shrinkage and Selection Operator) method, which simultaneously performs variable selection and regularization to enhance the predictive accuracy and interpretability of statistical models. This method is particularly useful in high-dimensional data settings, such as genomics and proteomics, where the number of variables often exceeds the number of observations. Zou and Hastie (2005) further extended this approach with the Elastic Net, which combines the penalties of the Lasso and Ridge regression, offering a more flexible tool for variable selection in complex datasets.

The need for robust and interpretable clinical prediction models has led to the development of advanced model-building strategies. Alonzo (2009) discusses the principles and applications of clinical prediction models, emphasizing the importance of model validation and updating in clinical research. These models are used to predict patient outcomes based on individual characteristics and clinical variables, providing valuable tools for personalized medicine. The integration of advanced statistical techniques, such as penalized regression and machine learning, into these models enhances their predictive performance and clinical utility (Greenwood et al., 2020: Jamshidi, Pelletier & Martel-Pelletier, 2019).

Mediation analysis is another innovative methodology that has gained prominence in clinical research, particularly in understanding the mechanisms through which interventions exert their effects. Böhnke (2016) offers an in-depth examination of mediation and interaction methods, which are crucial for understanding the pathways through which treatments affect outcomes. These techniques enable researchers to separate the direct and indirect effects of interventions, offering more profound insights into the mechanisms at play in clinical studies.

Fractional polynomials have emerged as a powerful tool for modeling continuous variables in clinical research. Royston and Sauerbrei (2008) advocate for the use of fractional polynomials in multivariable model-building, offering a pragmatic approach to handling non-linear relationships between continuous predictors and outcomes. This methodology enhances the flexibility of regression models, allowing for more accurate representation of complex relationships in clinical data.

The Cox proportional hazards model, introduced by Cox (1972), remains one of the most widely used methods for analyzing time-to-event data in clinical research. This model assumes that the hazard ratio between different levels of a covariate is constant over time, an assumption that may not always hold in practice. However, innovations such as

time-dependent covariates and stratified Cox models have been developed to address these limitations, allowing for more accurate modeling of survival data in clinical studies.

Finally, the integration of machine learning techniques into clinical research has opened new avenues for analyzing and interpreting complex datasets. These techniques, which include random forests, support vector machines, and neural networks, offer powerful tools for identifying patterns and relationships in large-scale clinical data. While traditional statistical methods remain essential, the incorporation of machine learning allows for the analysis of data in ways that were previously not possible, enhancing the ability

2.3. Case Studies and Practical Applications

The application of advanced statistical methods in clinical research has led to significant advancements in the accuracy, reliability and interpretability of study findings. These methods are not merely theoretical constructs but have been applied successfully in various case studies, demonstrating their practical utility in addressing complex clinical questions. For instance, innovative statistical methodologies have been crucial in the transcriptomic analysis and mass balance studies of polycyclic aromatic hydrocarbons (PAHs) in environmental samples, where these methods facilitated the identification of key genes and metabolic pathways involved in the degradation process. Such analyses have significantly contributed to understanding the biodegradation mechanisms at a molecular level, thereby providing insights into potential therapeutic applications and environmental remediation strategies. The study by Eregie et al. (2024) exemplifies this approach, highlighting the role of advanced statistical techniques in elucidating complex biological processes and guiding future research in both clinical and environmental contexts.

One of the most prominent applications of advanced statistical methods is in the development of clinical prediction models. Alonzo (2009) presents a comprehensive approach to developing and validating clinical prediction models, which are essential tools for predicting patient outcomes based on individual characteristics. These models are built using regression techniques that incorporate multiple variables, allowing clinicians to make informed decisions about patient care. For instance, in cardiology, prediction models are used to estimate the risk of adverse events such as heart attacks or strokes, enabling personalized treatment strategies that improve patient outcomes.

Another critical application is in the field of survival analysis, where the Cox proportional hazards model has been widely used to analyze time-to-event data (Cox, 1972). This model has been applied in numerous clinical studies, including cancer research, where it helps to estimate the effect of treatment on patient survival. For example, Royston and Parmar (2002) extended the Cox model to include flexible parametric approaches, allowing for a more nuanced analysis of survival data. This extension has been particularly useful in cases where the proportional hazards assumption does not hold, providing a more accurate assessment of treatment effects over time.

The use of causal inference methods has also been crucial in addressing the limitations of observational studies, where randomization is not possible. Stensrud et al. (2022) discuss the application of causal inference techniques such as inverse probability weighting and marginal structural models to estimate causal effects in the presence of time-varying confounders. These methods have been applied in studies evaluating the long-term effects of medications on chronic conditions, providing insights that would not be possible through traditional statistical methods alone. For instance, in the study of the effects of antiretroviral therapy on HIV patients, causal inference methods have helped to clarify the long-term benefits of treatment despite the presence of confounding factors.

Regularization techniques such as Lasso and Elastic Net have become invaluable in high-dimensional data settings, where the number of predictors exceeds the number of observations. Tibshirani (1996) introduced the Lasso method, which has been applied in genomic studies to identify biomarkers associated with disease outcomes. For example, in cancer research, Lasso has been used to select a subset of genes that are most predictive of patient survival, enabling the development of more targeted therapies. Zou and Hastie (2005) further enhanced this approach with the Elastic Net, which combines the strengths of Lasso and Ridge regression to handle correlated predictors, offering more robust variable selection in complex datasets.

Multiple imputation techniques have revolutionized the handling of missing data in clinical research. Rubin (2004) introduced the concept of multiple imputation, which has since become a standard practice in dealing with incomplete datasets. This method has been particularly impactful in longitudinal studies, where missing data can significantly bias results. For instance, in a study on the progression of Alzheimer's disease, multiple imputation was used to account for missing cognitive test scores, ensuring that the final analysis was not biased by incomplete data. White, Royston and Wood (2011) provided further guidance on the application of multiple imputation using chained equations, which has been widely adopted in clinical research to improve the robustness of statistical inferences.

The application of fractional polynomials in multivariable model-building has also had a significant impact on clinical research. Royston and Sauerbrei (2008) demonstrated the utility of fractional polynomials in modeling non-linear relationships between continuous predictors and outcomes. This approach has been applied in various fields, including epidemiology, where it helps to accurately model the dose-response relationship between exposure to risk factors and the development of diseases. For example, in studies on the effects of smoking on lung cancer risk, fractional polynomials have provided a more precise estimation of the risk at different levels of exposure, leading to more effective public health interventions.

Stratified randomization and dynamic balancing methods have been used to enhance the validity of clinical trials by ensuring that important covariates are balanced across treatment groups (O`Brien-Carelli et al, 2022). Signorini et al. (1903) discuss the application of these methods in cancer trials, where they have been used to ensure that patient characteristics such as age, gender, and disease stage are evenly distributed between treatment arms. This approach reduces the potential for confounding and enhances the credibility of the trial results. In a landmark breast cancer trial, stratified randomization was used to ensure that key prognostic factors were balanced, leading to more reliable conclusions about the efficacy of the treatment under investigation.

Mediation analysis has been applied to understand the mechanisms through which interventions exert their effects. Böhnke (2016) outlines a framework for mediation analysis, which has been applied in behavioral health research to investigate the mechanisms through which psychological interventions affect health outcomes, such as stress reduction or enhanced coping skills. For instance, in a study examining the effectiveness of cognitive-behavioral therapy (CBT) for depression, mediation analysis revealed that improvements in depressive symptoms were mediated by alterations in negative thinking patterns, offering insights into the psychological processes underlying the efficacy of CBT.

In the field of oncology, the application of advanced statistical methods has led to significant improvements in the analysis and interpretation of clinical trial data. For example, in a large-scale trial evaluating the effectiveness of a new chemotherapy regimen, advanced survival analysis techniques were used to account for time-dependent covariates and non-proportional hazards, leading to more accurate estimates of treatment efficacy. The use of these innovative statistical methods has provided oncologists with more reliable evidence to guide treatment decisions, ultimately improving patient outcomes.

Finally, the integration of machine learning techniques with traditional statistical methods has opened new avenues for analyzing complex clinical datasets. These techniques, including random forests and neural networks, have been applied in predictive modeling, image analysis, and personalized medicine. For instance, in a study on predicting patient outcomes in intensive care units (ICUs), machine learning models were combined with traditional regression techniques to identify the most important predictors of mortality, leading to the development of more effective clinical decision support systems.

These case studies and practical applications demonstrate the power and versatility of advanced statistical methodologies in addressing the challenges of modern clinical research. By leveraging these innovative approaches, researchers can derive more accurate, reliable, and clinically relevant insights from their data, ultimately contributing to the advancement of medical science and the improvement of patient care.

2.4. Future Directions in Biostatistics and Clinical Research

The future of biostatistics and clinical research is poised for significant evolution as emerging technologies, advanced methodologies, and interdisciplinary approaches continue to reshape the landscape of biomedical science. The integration of novel statistical techniques, the increasing role of big data and machine learning, and the need for more sophisticated causal inference models are among the key areas that will define the future direction of this field. As clinical research becomes more complex and data-driven, the role of biostatistics will be increasingly critical in ensuring the validity, reliability, and applicability of research findings (Stensrud et al., 2020). Additionally, the integration of biostatistics with environmental studies, such as the comparative analysis of biodegradation processes involving Scenedesmus vacuolatus versus microalgal consortia, further highlights the expanding scope of biostatistics in addressing complex, interdisciplinary challenges (Eregie & Jamal-Ally, 2019).

One of the most promising future directions in biostatistics is the further development and application of advanced regression models. Harrell (2012) emphasizes the importance of regression modeling strategies in clinical research, particularly in handling complex datasets with multiple variables. The future will likely see greater use of non-linear and non-parametric regression techniques, such as machine learning algorithms, that can handle large, high-dimensional data. These techniques will provide more accurate predictions and deeper insights into the relationships

between variables, enabling more personalized and precise treatment strategies in clinical practice (Hazra & Gogtay, 2017; Pisică et al., 2022).

The shift away from traditional reliance on p-values and towards more comprehensive statistical inference methods will also continue to gain traction. McShane et al. (2019) advocate for abandoning the strict use of p-values as a measure of statistical significance, arguing that this approach often oversimplifies complex data and leads to misinterpretation of results. The future of biostatistics will likely involve a greater emphasis on effect sizes, confidence intervals, and Bayesian methods, which provide a more nuanced understanding of the data. This shift will enhance the interpretability of clinical research findings and ensure that statistical analyses are more aligned with the practical and clinical significance of the results.

Causal inference will remain a central focus in the future of biostatistics, particularly as researchers continue to grapple with the challenges of drawing valid conclusions from observational data (O`Brien-Carelli et al, 2022). Stensrud et al. (2022) discuss the potential of advanced causal inference methods, such as marginal structural models and g-methods, to address the limitations of traditional statistical approaches. The future will likely see the development of more sophisticated causal inference techniques that can better account for time-varying confounders and other complexities inherent in real-world data. These advancements will be crucial for ensuring that causal claims in clinical research are both robust and credible.

The role of machine learning and artificial intelligence (AI) in biostatistics is another area that is expected to expand significantly in the coming years. Machine learning techniques, such as random forests, support vector machines, and deep learning, have already shown promise in analyzing large-scale clinical datasets and identifying complex patterns that traditional statistical methods might miss. Tibshirani (1996) and Zou and Hastie (2005) underscore the effectiveness of regularization techniques such as Lasso and Elastic Net in managing high-dimensional datasets, a challenge that has become increasingly significant in the big data era. Looking ahead, these methods are expected to be integrated with machine learning algorithms, further improving predictive modeling and variable selection in clinical research (James et al., 2013).

The application of big data analytics in biostatistics will also play a significant role in shaping the future of clinical research. With the increasing availability of large-scale datasets from electronic health records (EHRs), genomics, and other sources, there is a growing need for advanced statistical methods that can efficiently process and analyze these vast amounts of data. Alonzo (2009) discusses the development of clinical prediction models that can leverage big data to improve patient outcomes. The future will likely see the integration of big data analytics with traditional biostatistical methods, enabling more comprehensive analyses that can uncover new insights into disease mechanisms and treatment effects.

The handling of missing data remains a critical challenge in clinical research and future developments in this area will likely focus on improving the robustness and accuracy of imputation methods. Rubin (2004) introduced multiple imputation as a method for dealing with missing data, and this approach has since become a standard in the field. However, as datasets become larger and more complex, there will be a need for more advanced imputation techniques that can handle the intricacies of modern clinical research data. Enders (2017) and White, Royston and Wood (2011) offer detailed guidance on applying multiple imputation through chained equations. However, future advancements may see the incorporation of machine learning techniques to improve the precision and dependability of imputed data.

The development of new statistical models for analyzing survival data is another area of focus for the future of biostatistics. The Cox proportional hazards model, introduced by Cox (1972), remains one of the most widely used methods for analyzing time-to-event data. However, there is a growing recognition that the assumptions underlying this model may not always hold in practice, particularly in the presence of time-dependent covariates or non-proportional hazards. Future developments in survival analysis will likely involve the creation of more flexible models that can better account for these complexities, such as those proposed by Royston and Sauerbrei (2008) in their work on fractional polynomials and by Royston and Parmar (2002) in their development of flexible parametric survival models.

For instance, big data analytics has enabled healthcare providers to analyze large datasets to identify trends, predict outcomes, and personalize treatments (Ogundipe & Oghenetejiri, 2024). The importance of interdisciplinary collaboration in biostatistics and clinical research cannot be overstated, and future developments will likely emphasize the need for closer integration between statisticians, clinicians, and data scientists. Böhnke (2016) explores the role of mediation analysis in elucidating the mechanisms by which interventions achieve their effects, a process that frequently demands multidisciplinary collaboration. As clinical research grows increasingly complex and data-driven, integrating expertise from various fields will be crucial to ensuring that statistical analyses remain both robust and clinically

applicable. The integration of innovative methodologies, such as those demonstrated in the synergistic effect of process parameters and nanoparticles on the biodegradation of spent lubricant oil waste by UV-exposed Scenedesmus vacuolatus, highlights the necessity of combining expertise from different scientific domains to achieve optimal outcomes in research (Eregie et al., 2023).

Finally, the ethical considerations surrounding the use of advanced statistical methods in clinical research will become increasingly important in the future. As biostatistics continues to evolve, there will be a need for more robust frameworks to ensure that statistical analyses are conducted in a way that is both scientifically rigorous and ethically sound. This includes addressing issues such as data privacy, the potential for algorithmic bias in machine learning models, and the need for transparency in statistical reporting. Recent legislation, such as the Nigeria Data Protection Act, has established principles and obligations that significantly impact digital records management practices in healthcare and other sectors (Ogundipe, 2024). The future of biostatistics will likely involve the development of new guidelines and standards that address these ethical concerns, ensuring that the field continues to contribute to the advancement of medical science in a responsible and ethical manner.

3. Conclusion

This study has successfully achieved its aim of exploring and addressing the challenges associated with the application of statistical methods in clinical research, while also providing innovative solutions and methodologies to enhance the accuracy and reliability of research findings. Through a detailed examination of fundamental concepts in biostatistics, the study identified the complexities and limitations inherent in traditional statistical approaches, particularly in the context of modern clinical trials and observational studies.

Key findings from the study highlighted the critical role of advanced regression models, causal inference techniques, and machine learning algorithms in overcoming the limitations of conventional statistical methods. The integration of these methodologies into clinical research has proven to be essential for handling the intricacies of high-dimensional data, accounting for confounding variables, and improving the overall validity of research outcomes. The study also underscored the importance of shifting away from strict reliance on p-values, advocating for a more nuanced approach that considers effect sizes, confidence intervals, and practical significance.

In conclusion, this study has demonstrated that by embracing innovative statistical solutions, researchers can significantly enhance the rigor and applicability of clinical research. The development of advanced prediction models, robust imputation techniques for missing data, and flexible survival analysis models were shown to be pivotal in improving the precision and interpretability of research findings. Additionally, the study emphasized the need for interdisciplinary collaboration and the ethical application of these advanced methodologies, ensuring that biostatistics continues to contribute meaningfully to the advancement of medical science.

Recommendations from this study include the continued development and adoption of advanced statistical techniques in clinical research, particularly those that integrate machine learning and big data analytics. Furthermore, researchers are encouraged to move beyond traditional statistical paradigms and adopt more comprehensive approaches that consider the broader context and clinical relevance of their findings. By doing so, the field of biostatistics will be well-positioned to meet the challenges of modern clinical research and contribute to the ongoing improvement of patient care and treatment outcomes.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Alonzo, T.A., (2009). Clinical prediction models: a practical approach to development, validation, and updating: by Ewout W. Steyerberg. https://doi.org/10.1093/aje/kwp129

[2] Armijo-Olivo, S., (2018). The importance of determining the clinical significance of research results in physical therapy clinical research. Brazilian journal of physical therapy, 22(3), p.175. doi: 10.1016/j.bjpt.2018.02.001.

[3] Bland, J.M. and Altman, D., (1986). Statistical methods for assessing agreement between two methods of clinical measurement. The lancet, 327(8476), pp.307-310. https://doi.org/10.1016/S0140-6736(86)90837-8

[4] Böhnke, J.R., (2016). Explanation in causal inference: methods for mediation and interaction. https://doi.org/10.1080/17470218.2015.1115884

[5] Casey, J.D., Beskow, L.M., Brown, J., Brown, S.M., Gayat, É., Gong, M.N., Harhay, M.O., Jaber, S., Jentzer, J.C., Laterre, P.F. and Marshall, J.C., (2022). Use of pragmatic and explanatory trial designs in acute care research: lessons from COVID-19. The Lancet Respiratory Medicine, 10(7), pp.700-714.

[6] Clark, T.G., Bradburn, M.J., Love, S.B. and Altman, D.G., (2003). Survival analysis part I: basic concepts and first analyses. British journal of cancer, 89(2), pp.232-238.

[7] Cox, D.R., (1972). Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2), pp.187-202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

[8] Davis, S.L., Johnson, A.H., Lynch, T., Gray, L., Pryor, E.R., Azuero, A., Soistmann, H.C., Phillips, S.R. and Rice, M., (2021). Inclusion of effect size measures and clinical relevance in research papers. Nursing research, 70(3), pp.222-230.

[9] Enders, C.K., (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. Behaviour research and therapy, 98, pp.4-18. https://doi.org/10.1016/j.brat.2016.11.008

[10] Eregie, S.B. and Jamal-Ally, S.F., (2023). Comparison of biodegradative efficiency of wildtype versus mutagenised Scenedesmus vacuolatus of spent coolant waste: dehydrogenase activity and total petroleum degradation studies. International Journal of Environmental Analytical Chemistry, 103(18), pp.6996-7022. https://doi.org/10.1080/03067319.2021.1965593

[11] Eregie, S.B. and Jamal-Ally, S.F., (2019). Comparison of biodegradation of lubricant wastes by Scenedesmus vacuolatus vs a microalgal consortium. Bioremediation Journal, 23(4), pp.277-301. https://doi.org/10.1080/10889868.2019.1671792

[12] Eregie, S.B., Sanusi, I.A., Kana, G.E. and Ademola, O.O., (2023). Synergistic effect of process parameters and nanoparticles on spent lubricant oil waste biodegradation by UV-exposed Scenedesmus vacuolatus: Process modelling, kinetics and degradation pathways. Bioresource Technology Reports, 24, p.101627. https://doi.org/10.1016/j.biteb.2023.101627

[13] Eregie, S.B., Sanusi, I.A., Kana, G.E. and Olaniran, A.O., (2024). Effect of ultra-violet light radiation on Scenedesmus vacuolatus growth kinetics, metabolic performance, and preliminary biodegradation study. Biodegradation, 35(1), pp.71-86. https://doi.org/10.1007/s10532-023-10029-2

[14] Eregie, S.B., Sanusi, I.A., Kumar, A., Kana, G.E. and Ademola, O.O., (2024). Transcriptomic removal and mass balance of polycyclic aromatic hydrocarbons in waste spent coolant oil: Gene discovery, enzyme identification and metabolic pathway. Bioresource Technology Reports, 27, p.101908. https://doi.org/10.1016/j.biteb.2024.101908

[15] Fahr, P., Buchanan, J. and Wordsworth, S., (2019). A review of the challenges of using biomedical big data for economic evaluations of precision medicine. Applied health economics and health policy, 17(4), pp.443-452. https://doi.org/10.1007/s40258-019-00474-7

[16] Fleming, T.R. and Harrington, D.P., (2013). Counting processes and survival analysis (Vol. 625). John Wiley & Sons.

[17] Freedman, L.S., Lowe, D. and Macaskill, P., (1984). Stopping rules for clinical trials incorporating clinical opinion. Biometrics, pp.575-586.

[18] Greenwood, C.J., Youssef, G.J., Letcher, P., Macdonald, J.A., Hagg, L.J., Sanson, A., Mcintosh, J., Hutchinson, D.M., Toumbourou, J.W., Fuller-Tyszkiewicz, M. and Olsson, C.A., (2020). A comparison of penalised regression methods for informing the selection of predictive markers. PloS one, 15(11), p.e0242730. https://doi.org/10.1371/journal.pone.0242730

[19] Harrell, F.E., (2001). Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis (Vol. 608). New York: springer.

[20] Harrell, F.E., (2012). Regression modeling strategies. R package version, pp.6-2.

[21] Hayden, J.A., van der Windt, D.A., Cartwright, J.L., Côté, P. and Bombardier, C., (2013). Assessing bias in studies of prognostic factors. Annals of internal medicine, 158(4), pp.280-286.

[22] Hazra, A. and Gogtay, N., (2017). Biostatistics series module 10: Brief overview of multivariate methods. Indian journal of dermatology, 62(4), pp.358-366.

[23] Jahun, I., Greby, S. M., Adesina, T., Agbakwuru, C., Dalhatu, I., Yakubu, A., Jelpe, T., Okoye, M., Ikpe, S., Ehoche, A., Abimiku, A., Aliyu, G., Charurat, M., Greenwell, G., Bronson, M., Patel, H., McCracken, S., Voetsch, A. C., Parekh, B., Swaminathan, M., … Aliyu, S. (2021). Lessons From Rapid Field Implementation of an HIV Population-Based Survey in Nigeria, 2018. Journal of acquired immune deficiency syndromes (1999), 87(Suppl 1), S36–S42. https://doi.org/10.1097/QAI.0000000000002709

[24] James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

[25] Jamshidi, A., Pelletier, J.P. and Martel-Pelletier, J., (2019). Machine-learning-based patient-specific prediction models for knee osteoarthritis. Nature Reviews Rheumatology, 15(1), pp.49-60.

[26] Joseph, O.B. & Uzondu, N.C., (2024). Bridging the Digital Divide in STEM Education: Strategies and Best Practices. Engineering Science & Technology Journal, 5(8), pp.2435-2453.

[27] Joseph, O.B. & Uzondu, N.C., (2024). Curriculums Development for Interdisciplinary STEM Education: A Review of Models and Approaches. International Journal of Applied Research in Social Sciences, 6(8), pp.1575-1592.

[28] Joseph, O.B. & Uzondu, N.C., (2024). Integrating AI and Machine Learning in STEM Education: Challenges and Opportunities. Computer Science & IT Research Journal, 5(8), pp.1732-1750.

[29] Joseph, O.B. & Uzondu, N.C., (2024). Professional Development for STEM Educators: Enhancing Teaching Effectiveness through Continuous Learning. International Journal of Applied Research in Social Sciences, 6(8), pp.1557-1574.

[30] Kang, M., Ragan, B.G. and Park, J.H., (2008). Issues in outcomes research: an overview of randomization techniques for clinical trials. Journal of athletic training, 43(2), pp.215-221. https://doi.org/10.4085/1062-6050-43.2.215

[31] Layode, O., Naiho, H., Adeleke, G., Udeh, E., & Talabi, T. (2024a). The Role of Cybersecurity in Facilitating Sustainable Healthcare Solutions: Overcoming Challenges to Protect Sensitive Data. International Medical Science Research Journal, 4(6), 668–693. https://doi.org/10.51594/imsrj.v4i6.1228.

[32] Layode, O., Naiho, H., Adeleke, G., Udeh, E., & Talabi, T. (2024b). Data Privacy and Security Challenges in Environmental Research: Approaches to Safeguarding Sensitive Information. International Journal of Applied Research in Social Sciences, 6(6), 1193–1214. https://doi.org/10.51594/ijarss.v6i6.1210.

[33] Lee, K.J., Moreno-Betancur, M., Kasza, J., Marschner, I.C., Barnett, A.G. and Carlin, J.B., (2019). Biostatistics: a fundamental discipline at the core of modern health data science. The Medical Journal of Australia, 211(10), p.444.

[34] Martínez-Mesa, J., González-Chica, D.A., Duquia, R.P., Bonamigo, R.R. and Bastos, J.L., (2016). Sampling: how to select participants in my research study? Anais brasileiros de dermatologia, 91(3), pp.326-330. https://doi.org/10.1590/abd1806-4841.20165254

[35] McShane, B.B., Gal, D., Gelman, A., Robert, C. and Tackett, J.L., (2019). Abandon statistical significance. The American Statistician, 73(sup1), pp.235-245. https://doi.org/10.1080/00031305.2018.1527253

[36] Nigeria HIV-AIDS Indicator and Impact Survey (NAIIS) 2018. NGA-FMOH-NAIIS-2018-v2.1. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://naiis.ng/resource/NAIIS-Report-2018.pdf

[37] O'Brien-Carelli, C., Steuben, K., Stafford, K. A., Aliogo, R., Alagi, M., Johanns, C. K., Ibrahim, J., Shiraishi, R., Ehoche, A., Greby, S., Dirlikov, E., Ibrahim, D., Bronson, M., Aliyu, G., Aliyu, S., Dwyer-Lindgren, L., Swaminathan, M., Duber, H. C., & Charurat, M. (2022). Mapping HIV prevalence in Nigeria using small area estimates to develop a targeted HIV intervention strategy. *PloS one*, *17*(6), e0268892. https://doi.org/10.1371/journal.pone.0268892

[38] Ogundipe, O., & Aweto, T. (2024). The shaky foundation of global technology: A case study of the 2024 crowdstrike outage. International Journal of Multidisciplinary Research and Growth Evaluation, 5(5), 106-108.

[39] Ogundipe, G., & Oghenetejiri, D. (2024). Supply Chain Analysis of the Global Fund Malaria Elimination Project in Nigeria. Journal of Biodivers. Conservation, 8(3), 97-112.

[40] Ogundipe, O. A. (2024). Managing digital records within Nigeria's regulatory framework. International Journal of Science and Research Archive, 12(02), 2861–2868

[41] Pannucci, C.J. and Wilkins, E.G., (2010). Identifying and avoiding bias in research. Plastic and reconstructive surgery, 126(2), pp.619-625.

[42] Patsopoulos, N.A., (2011). A pragmatic view on pragmatic trials. Dialogues in clinical neuroscience, 13(2), pp.217-224. https://doi.org/10.31887/DCNS.2011

[43] Perera, M. and Dwivedi, A.K., (2020). Statistical issues and methods in designing and analyzing survival studies. Cancer Reports, 3(4), p.e1176. https://doi.org/10.1002/cnr2.1176

[44] Pisică, Dana, Ruben Dammers, Eric Boersma, and Victor Volovici. "Tenets of good practice in regression analysis. a brief tutorial." World neurosurgery 161 (2022): 230-239. https://doi.org/10.1016/j.wneu.2022.02.112

[45] Pocock, S.J. and Simon, R., (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics, pp.103-115.

[46] Royston, P. and Parmar, M.K., (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in medicine, 21(15), pp.2175-2197. https://doi.org/10.1002/sim.1203

[47] Royston, P. and Sauerbrei, W., 2008). Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables. John Wiley & Sons.

[48] Rubin, D.B., (2004). Multiple imputation for nonresponse in surveys (Vol. 81). John Wiley & Sons.

[49] Signorini, D.F., Leung, O., Simes, R.J., Beller, E., Gebski, V.J. and Callaghan, T., (1993). Dynamic balanced randomization for clinical trials. Statistics in medicine, 12(24), pp.2343-2350. https://doi.org/10.1002/sim.4780122410

[50] Schober, P. and Vetter, T.R., (2018). Survival analysis and interpretation of time-to-event data: the tortoise and the hare. Anesthesia & Analgesia, 127(3), pp.792-798.

[51] Schwartz, D. and Lellouch, J., (1967). Explanatory and pragmatic attitudes in therapeutical trials. Journal of chronic diseases, 20(8), pp.637-648. https://doi.org/10.1016/0021-9681(67)90041-0

[52]

[53] Serdar, C.C., Cihan, M., Yücel, D. and Serdar, M.A., (2021). Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochemia medica, 31(1), pp.27-53. https://doi.org/10.11613/BM.2021.010502

[54] Seyi- Lande, O., Layode, O., Naiho, H., Adeleke, G., Udeh, E., & Talabi, T. (2024). Circular Economy and Cybersecurity: Safeguarding Information and Resources in Sustainable Business Models. Finance & Accounting Research Journal, 6(6), 953–977. https://doi.org/10.51594/farj.v6i6.1214.

[55] Sloan, J.A. and Dueck, A., (2004). Issues for statisticians in conducting analyses and translating results for quality of life end points in clinical trials. Journal of biopharmaceutical statistics, 14(1), pp.73-96. https://doi.org/10.1081/BIP-120028507

[56] Stensrud, M.J., Young, J.G., Didelez, V., Robins, J.M. and Hernán, M.A., (2022). Separable effects for causal inference in the presence of competing events. Journal of the American Statistical Association, 117(537), pp.175-183. https://doi.org/10.1080/01621459.2020.1765783

[57] Tashkin, D.P., Amin, A.N. and Kerwin, E.M., (2020). Comparing randomized controlled trials and real-world studies in chronic obstructive pulmonary disease pharmacotherapy. International journal of chronic obstructive pulmonary disease, pp.1225-1243.

[58] Tibshirani, R., (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), pp.267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[59] White, I.R., Royston, P. and Wood, A.M., (2011). Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine, 30(4), pp.377-399. https://doi.org/10.1002/sim.4067

[60] Zou, H. and Hastie, T., (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2), pp.301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x