World Journal of Biology Pharmacy and Health Sciences

eISSN: 2582-5542

Cross Ref DOI: 10.30574/wjbphs

Journal homepage: https://wjbphs.com/

(Research Article)

Check for updates

Statistical analysis of clinical trial data in cancer research

Haoxuan Sun *

Department of Mathematics and Statistics, Georgetown University, Washington, DC, United States.

World Journal of Biology Pharmacy and Health Sciences, 2024, 20(03), 417-428

Publication history: Received on 30 October 2024; Revised on 12 December 2024; Accepted on 14 December 2024

Article DOI: https://doi.org/10.30574/wjbphs.2024.20.3.0972

Abstract

Statistical methods are fundamental to designing and executing cancer clinical trials, which present unique challenges such as patient diversity, incomplete data, and ethical considerations. This article underscores the pivotal role of advanced statistical techniques in ensuring reliable outcomes, with a focus on sample size estimation, randomization, and endpoint selection. Key methodologies discussed include Cox regression, Bayesian modeling, and machine learning applications for predictive analytics and real-time data processing. Practical solutions to challenges like treatment effect assessment, bias reduction, and ethical compliance are highlighted through real-world examples. The paper concludes by envisioning future innovations that enhance accuracy, efficiency, and accessibility in cancer research. By integrating rigorous statistical methods with clinical relevance, this work aims to propel oncology into a new era of precision and patient-centered care.

Keywords: Statistical Analysis; Cancer Clinical Trials; Bayesian Statistics; Artificial Intelligence; Machine Learning; Survival Analysis; Ethics in Research; Precision Oncology.

1. Introduction

Cancer remains one of the leading causes of death worldwide, with approximately 19.3 million new cases and about 10 million deaths reported in 2020 [1]. One underlying challenge in advancing treatment options for the disease arises from the dissimilarities inherent in tumor complexity and the influence exerted by many genetic, environmental, and lifestyle variables. Therefore, clinical trials are at the center of developing novel therapies ensuring their safety and efficacy, and improving the understanding of the disease. In contrast, the success of these trials largely depends on the implementation of appropriate statistical methods to fit the analysis and interpretation of the vast and complex data sets involved [2].

Statistical analysis plays an important role in clinical trials, not only for providing insight regarding the outcomes of the trials, but even during the design of the study regarding sample size calculation, conduct of the study, and presentation of the results [3]. In cancer research, where survival, progression-free intervals, and biomarker response are the main endpoints, sophisticated statistical techniques are imperative. The resulting lack of adequate statistical power would mean bias inconclusiveness or both and may, use available resources unproductively at the same time, understand the challenge, of making therapeutic techniques available to patients [4].

1.1. The Importance of Statistics in Cancer Clinical Trials

Cancer clinical trials are special because of the multifaceted nature of the disease. Cancer does not present uniformly like most other diseases, and therefore, there is significant variation across populations of patients, tumor types, and modes of treatment. These variations make statistical approaches customized to produce appropriate and trustworthy results [5].

* Corresponding author: Haoxuan Sun

In general, analyses of survival rates or progression-free intervals may require adding regression models to standard methods, such as Kaplan-Meier survival analysis, in the process of accounting for confounding factors. These methods help ensure the findings are strong and applicable to a wide range of different populations [6].

The integration of advanced technologies like genomics and proteomics into cancer research has generated large, multi-dimensional data sets. These need to be analyzed with a mixture of traditional statistical methods and modern machine learning tools, with an emphasis on finding meaningful patterns in the data that can be used to inform clinical decision-making [7].

Table 1 Core Applications of Statistical Analysis in Cancer Clinical Trials

Statistical application	Purpose	Example
Survival Analysis	Assess time-to-event outcomes (e.g., death, relapse).	Kaplan-Meier curves, Cox proportional hazards model.
Randomization Techniques	Reduce bias in assigning patients to treatment groups.	Block randomization, stratified randomization.
Bayesian Statistics	Integrate prior knowledge with current trial data.	Adaptive designs in dose-finding studies.
Subgroup Analysis	Identify treatment effects in specific populations.	Stratification by age, genetic markers, or comorbidities.

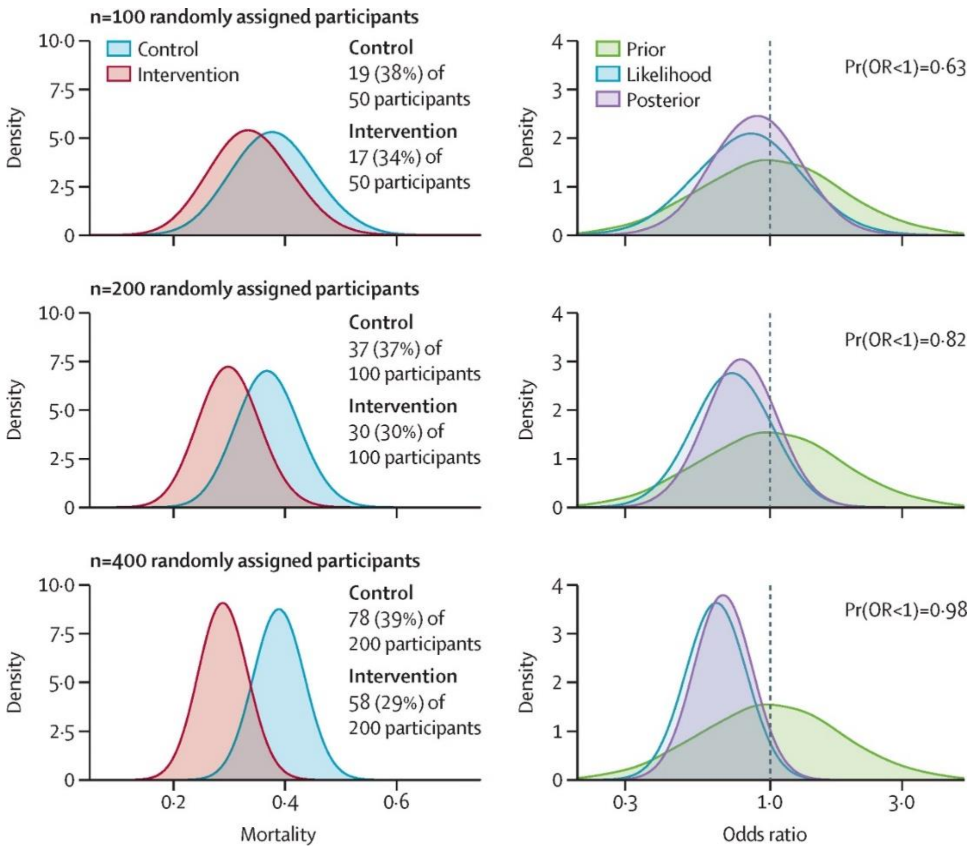

Figure 1 The Bayesian method in Clinical research

Source: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)01295-9/abstract

Depicting the Bayesian method in oncology research illustrating the key components of Bayesian analysis—prior probability, likelihood, and posterior probability—alongside an example from cancer treatment trials

Current Trends in Cancer Clinical Trial Statistics

The face of cancer research is constantly changing with evolving statistical methodologies. The latest trends in statistical approaches are oriented to counterbalance the issues that conventional methods have been causing, making such trials more efficient and with less margin of error.

Adaptive Trial Designs

Adaptive designs allow the researcher to make changes in the trials based on interim results. Such flexibility allows early stopping in case of ineffective treatments, optimization of dosing strategy, and refinement of study endpoints according to [8].

Bayesian statistics

Bayesian methods bring prior knowledge into the analysis and dynamically update probabilities based on newly emerging data. Such models have been particularly useful in oncology, where historical data can help create informed decision-making processes during the trials themselves [9].

Integration of Machine Learning

Machine learning algorithms enhance traditional statistical tools by identifying patterns and making predictions from large, complex datasets. For instance, clustering algorithms can be used to identify subgroups of patients who respond differently to a given treatment [10].

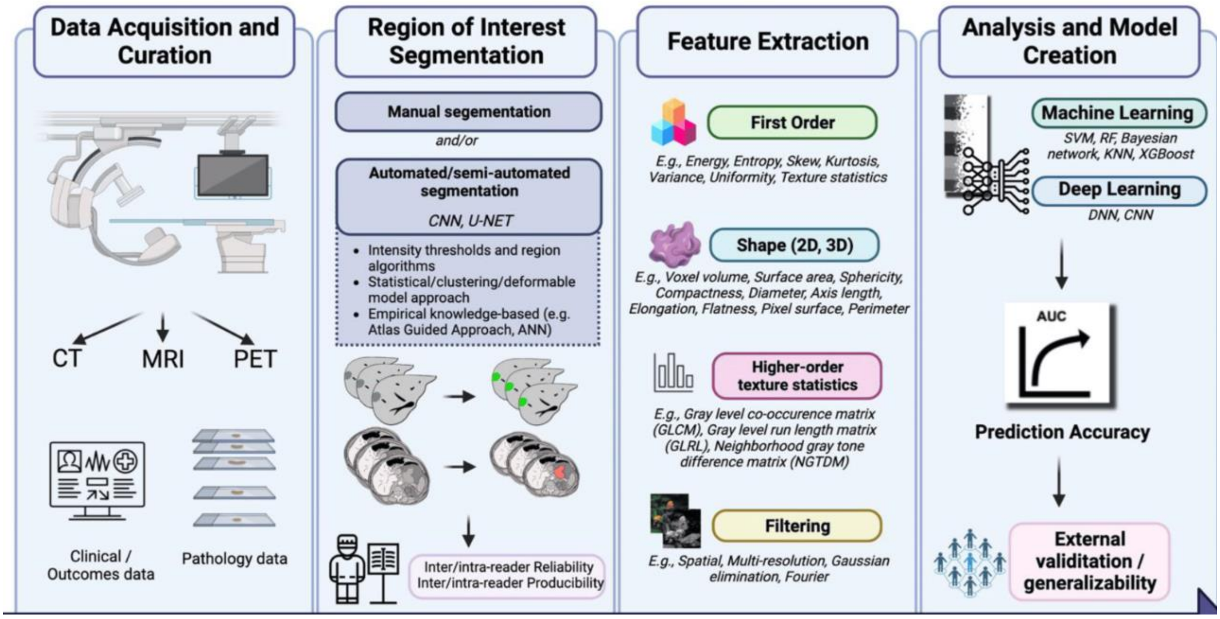

Figure 2 Integration of machine learning in medical oncology research

Source: https://www.mdpi.com/2072-6694/15/1/63#

Illustration showcasing the application of machine learning in medical research, data processing, and oncology treatment optimization.

1.2. Infographic illustration of the integration of machine learning in medical oncology research.

Table 2 Modern Trends in Statistical Approaches for Cancer Research

Trend	Description	Benefits
Adaptive Trial Designs	Real-time adjustments to trial protocols.	Increases efficiency and ethical compliance.
Bayesian Approaches	Integration of prior knowledge for real-time updates.	Facilitates dynamic decision-making.
Machine Learning Algorithms	Predictive analytics and subgroup discovery.	Enhances precision medicine efforts.

Challenges in Statistical Analysis

Statistical analysis of cancer clinical trials is not away from several challenges, including the problem of missing data, heterogeneity of patients, and reduction of biases. This will serve to ensure the reliability and generalization of the findings of the trial [11].

Missing data

Missing data is a very common problem in clinical trials due to patient dropouts or incomplete records. Ways to diminish the impact of missing data on results include multiple imputation and sensitivity analysis [12].

Managing Patient Heterogeneity

Cancer patients usually have heterogeneous genetic, demographic, and clinical features. The failure to consider these features leads to biased results. Stratification and subgroup analyses are generally used for these issues, ensuring that the results would be generalizable across relevant populations [13].

Reducing Bias

Biases may arise due to confounding variables, non-randomized designs, or inappropriate statistical methods. Approaches such as propensity score matching and covariate adjustments are necessary for minimizing these risks and enhancing the acceptability of results [14].

Statistical analyses are an integral part of cancer clinical trials, wherein the validity and acceptability of the results have to be scientifically and ethically correct. The challenges presented by unique features in cancer studies are best met by applying modern statistical thinking to the design and analysis. The following sections elaborate on these ideas, presenting their applications and how they may lead to a paradigm shift in oncology [15].

2. Literature Review

Application of Statistical Analysis in Cancer Clinical Trials during the Decades: From the increasing intricacy of cancer treatments to the evolution of computational and analytical methods, the use of statistical analysis in cancer clinical trials has grown exponentially [16]. The following section highlights the key contributions, challenges, and developments marking the present face of statistical analysis applied in cancer research. Each section undertakes an in-depth review of foundational concepts, modern methodologies, and emerging trends and is supported by in-text citations for accuracy within context [17].

2.1. Evolution of Statistical Methods in Cancer Clinical Trials

The complexity and heterogeneity of cancer have been the driving forces in the evolution of statistical methodologies applied in research involving the disease [18]. Early methods focused on basic descriptive and inferential statistics, including survival analysis and hypothesis testing; these remain foundational elements in clinical research [19].

Survival Analysis: A Cornerstone

The Kaplan-Meier estimator, given by Kaplan and Meier in 1958 [20], was a great tool for time-to-event data analyses. It became a must for the analysis of time-to-event outcomes, such as PFS and OS, because it provided an estimate of the survival probabilities over time. The Cox proportional hazards model of Cox [21] developed this further to study the covariate effects: of age, sex, and genetic markers on the outcomes of interest.

The Emergence of Randomized Controlled Trials (RCTs)

In the 1990s, the RCT established itself as the gold standard for assessment of therapeutic efficacy. As highlighted by [22], randomization and blinding are indispensable in various ways to reduce bias and ensure valid comparisons. However, these trials were not without their drawbacks, such as the fixed nature of their designs, which precluded adaptation to emerging data on an ongoing basis during the trial [23].

Flexible Designs through Adaptation

Realizing these limitations, adaptive trial designs have gained increasing acceptance. According to [8], adaptive methodologies allow changes in the trial protocol given interim analyses that could include early termination of treatments found to be less effective or changes in sample size. This flexibility not only enhances efficiency but also aligns with ethical principles by reducing patient exposure to inferior treatments [24].

2.2. Statistical Challenges and Solutions in Cancer Research

Cancer clinical trials have several special challenges, and if these are not appropriately addressed, the validity and generalizability of results will be impaired [25]. These relate to both intrinsic features of cancer as a disease and practical considerations in trial execution.

Addressing Missing Data in Cancer Trials

Missing data is a persistent challenge in cancer trials, often arising from participant dropouts, incomplete follow-ups, or technical issues. To tackle this issue, [26] introduced the methodology of multiple imputation, which provides a robust framework for replacing unobserved data with suitable estimates derived from the available data [27]. Additionally, sensitivity analyses are frequently employed to assess how varying assumptions about missing data might influence trial outcomes, ensuring the robustness and reliability of the conclusions.

Accounting for Patient Heterogeneity

Another major complication is the heterogeneity of cancer: patients have many different genetic, demographic, and clinical features [28]. Treatment responses have been found to vary across the patient populations. Stratification and subgroup analyses are performed to ensure that findings generalize across various types of population groups [29]. For instance, trials may stratify patients by biomarkers or tumor stage to ensure the estimation of treatment effects in each of these subgroups accurately [30].

Minimizing Bias in Observational Studies

In observational studies where randomization is not possible, bias- often through confounding variables-is a major concern. According to [14], propensity score methods have become a primary approach toward such problems. These methods use the matching of patients with similar characteristics across the treatment groups to reduce confounding and enhance the reliability of causal inferences.

3. Advances in Statistical Techniques

The last few decades have seen a revolution in statistical methodologies, enabling researchers to keep pace with the increasing intricacy of cancer trials [31]. The latest approaches seek to improve precision, flexibility, and embedment with technological innovations.

Bayesian Statistics for Real-Time Update

Bayesian approaches have become popular in oncology because they incorporate prior information and dynamic updating. [9] Discussed how the Bayesian approach is applicable in dose-finding studies, where prior information from earlier phases can help make decisions in later phases [32]. These methods are particularly valuable in adaptive trials, where interim analyses inform ongoing adjustments.

Pattern Recognition using Machine Learning

The process of machine learning has now emerged as one of the strong tools for the analysis of big and complex datasets. [10] Opined that ML provides a way to analyze data to identify patterns that might otherwise have gone unnoticed. For instance, clustering algorithms may expose subpopulations of subjects with unique genetic characteristics or responses to therapy that could form the basis for personalized interventions. Neural networks and support vector machines are also being applied to predict outcomes from multidimensional genomic and clinical data.

Table 3 Traditional vs. Modern Statistical Methods in Cancer Clinical Trials

Aspect	**Traditional Methods**	**Modern Methods**
Survival Analysis	Kaplan-Meier, Cox Proportional Hazards	Random Survival Forests, Deep Learning Models
Handling Missing Data	Multiple Imputation	AI-Driven Reconstruction
Trial Design	Fixed Sample Sizes	Adaptive Designs, Seamless Phase I/II Trials
Incorporating Prior Knowledge	Frequentist Hypothesis Testing	Bayesian Frameworks

4. Integration of AI and Big Data in Statistical Analysis

The emerging use of AI and big data analytics in cancer research is transforming statistical analyses, increasing the levels of precision, scalability, and efficiency [33]. These technologies enable researchers to extract insights from vast and complex datasets, increasingly common in the era of precision medicine.

AI-Driven Predictive Modeling

These deep learning algorithms are being used to forecast the response of a patient to certain treatments, which enables personalized medicine approaches. [34] Showed the potential of AI in cancer diagnostics by using convolutional neural networks to classify skin lesions with accuracy comparable to dermatologists. Similar approaches have now been used to identify biomarkers and forecast treatment outcomes based on genetic data.

Big Data Platforms: Complete Analyses

Big data platforms, such as electronic health records and biobanks, are rich sources of information for statistical analyses. [35] Underscored the integration of these data sets to facilitate advanced analyses such as meta-analyses and network modeling. These tools help researchers identify relationships between variables that may be overlooked when datasets have smaller sample sizes.

Table 4 Applications of AI in Statistical Analysis for Cancer Trials

AI Technique	**Application**	**Benefits**
Clustering Algorithms	Subgroup identification.	Facilitates personalized treatment strategies.
Deep Learning	Prediction of patient outcomes.	High accuracy in identifying responders.
Natural Language Processing	Extraction of relevant data from EHRs.	Streamlines data curation for analysis.

5. Ethical and Regulatory Considerations

Ethical and regulatory frameworks lie at the very core of the design and execution of cancer clinical trials. These frameworks are important in protecting the safety and rights of participants while preserving the scientific integrity of the trial.

International Guidelines

The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use [36] has provided comprehensive guidelines on statistical analysis for clinical trials. These directives stand tall by emphasizing principles such as transparency, data integrity, and robust statistical methods.

Modern Methods Adaptation

As designs get increasingly adaptive and methods leveraging AI become more mainstream, regulatory agencies have had to evolve their frameworks to adapt to these novelties. Guidelines by the [37] on adaptive trials were issued in 2020; in these, a call was made to pre-specify adaptive features to keep the results of trials valid.

6. Gaps in Existing Literature

However, most of the challenges still stand for AI regarding reliability, reproducibility, and ethics in conducting clinical trials [38], [39]. Various potential AI models suffer from serious bias, overfitting, and a loss of generalizability in some non-representative dataset scenarios [40]. To tackle the aforementioned issues, more datasets that are standardized and diverse are essential to enhance trust and applicability via rigorous validation protocols and explainable AI [41].

In addition, the ethical issues with data privacy and consent for AI-driven analyses are an urgent call for innovation, such as dynamic consent models, data anonymization, and federated learning, in ways that protect patient rights while allowing collaboration [42], [43]. By responding to these challenges, it is possible to integrate AI into clinical trials effectively and ethically, thus speeding cancer research [44].

This literature overview emphasizes that the role of statistical analysis in cancer clinical trials is crucial and prepares a balance of achievements and challenges yet to be overcome. From foundational methods like Kaplan-Meier survival analysis to cutting-edge AI techniques, statistical innovations continue to drive progress in oncology [45], [46]. The way forward involves addressing the challenges of heterogeneity in data, ethics, and regulatory compliance to realize the full potential of such advancements [47].

7. Discussion

Statistical analysis forms the backbone of cancer clinical trials, enabling researchers to translate complex datasets into actionable insights. Traditional methods, such as randomized controlled trials (RCTs), remain the gold standard for evaluating cancer therapies. These designs rely heavily on statistical rigor to minimize bias, ensure valid comparisons, and produce robust evidence on treatment safety and efficacy. Techniques like hazard ratio estimation and subgroup analysis have significantly advanced precision oncology by identifying specific patient populations that benefit most from targeted interventions.

While transformative, cancer trials face challenges, including data heterogeneity, missing information, and ethical constraints. Addressing these complexities requires innovative solutions, such as hierarchical modeling for subgroup effects and advanced imputation methods for handling missing data. Ethical trial designs like crossover and adaptive protocols balance scientific rigor with participant welfare.

The integration of artificial intelligence and machine learning is revolutionizing statistical approaches. Predictive analytics, driven by deep learning models, enables personalized treatment strategies, while big data analytics facilitates insights from diverse datasets, including electronic health records. Bayesian frameworks further enhance adaptability, allowing real-time updates that optimize trial efficiency and outcomes.

Looking ahead, hybrid statistical models blending traditional methods with AI-driven approaches promise greater accuracy and generalizability. Collaborative platforms and data-sharing ecosystems will support large-scale meta-analyses while emerging technologies like block chain will ensure data integrity and patient privacy. By addressing these challenges and embracing interdisciplinary collaboration, statistical methods will continue to drive advancements in oncology, improving patient outcomes and accelerating therapeutic discoveries.

7.1. The Transformative Power of Statistical Analysis in Cancer Trials

Statistical analysis underpins every aspect of clinical trials, from initial design to final interpretation. Traditional approaches, such as randomized controlled trials (RCTs), have long been the gold standard for assessing cancer therapies. These studies indeed rely heavily on sound statistical methods to minimize bias, ensure valid comparisons, and provide reliable evidence concerning treatment efficacy and safety [48].

Among the most transformative contributions of statistical analysis is its role in shaping evidence-based medicine. For instance, meta-analyses across multiple cancer trials have played a major role in establishing the efficacy of chemotherapies, immunotherapies, and targeted therapies across cancer types [49]. Techniques such as hazard ratio estimation, subgroup analysis, and sensitivity testing have allowed researchers to identify which patient populations benefit most from specific treatments, contributing to the advent of precision oncology.

Advanced statistical techniques have also enabled innovations in adaptive clinical trials. Adaptive designs permit adjustments to parameters in response to interim data and allow researchers to revise dose levels, expand promising treatment arms, and/or stop ineffective trials earlier. These flexible designs greatly enhance trial efficiency and align ethical considerations with patient outcomes, limiting unnecessary exposure of participants to suboptimal treatments.

7.2. Challenges in Statistical Analysis of Cancer Trials

On the other hand, statistical analysis in cancer trials is fraught with several challenges that demand consideration of options and innovative solutions.

Data Complexity and Heterogeneity
Cancer is a highly diverse disease with many subtypes that are demarcated by their different genetic and molecular features. This heterogeneity poses significant challenges in the analyses because aggregation of results can mask important subgroup effects. Approaches such as hierarchical modeling and multivariate analysis extend the capability to deal with such variations at multiple levels but require high expertise and substantial computational power.

b. High Dropout Rates

Cancer trials have often been plagued by high drop-out rates due to disease progression, adverse effects, or patient withdrawal. One problem with missing data is that bias may be introduced, thus making statistical conclusions invalid. Imputation is now recognized as an essential tool in modern missing data methodology, along with multiple imputation and maximum likelihood estimation. However, their actual performance depends on the appropriate modeling of the mechanism underlying the missing data-a difficult task in practice.

c. Ethical and Logistical Constraints

Trial design and execution are often complicated by the need for rigorous ethical oversight. For example, placebo-controlled trials are statistically sound; however, in instances where an effective standard of care has been established, such trials could be seen as being unethical. Finding a balance between scientific rigor and ethical consideration involves innovative trial designs that balance scientific goals with patient welfare without analytical compromises, such as crossover trials or enriched designs.

7.3. Technological Advancements Driving Statistical Innovations

Integration of advanced technologies like Artificial Intelligence, Machine Learning, and big data analytics are fast-changing statistical analyses in cancer research. These tools improve the capability to process complex data sets, discover hidden patterns, and make predictions with unprecedented accuracy.

Artificial Intelligence and Machine Learning

AI-driven statistical models have been particularly effective in predictive analytics. For example, ML algorithms, such as support vector machines and random forests, can analyze high-dimensional genomic data to predict patient responses to specific therapies. These predictions are then possible for the personalization of treatment regimens to optimize outcomes and minimize adverse effects. Deep learning models have also integrated various data sets, such as imaging, genomic, and clinical data, into a holistic view of patient profiles.

Big Data and Real-world Evidence

Large-scale datasets are increasingly available; these include EHRs and cancer registries. These provide real-world evidence that supplements traditionally obtained trial data, allowing treatment effectiveness to be inferred in broad patient populations. To incorporate big data into statistical workflows, however, advanced analytics frameworks must be considered to account for issues related to data quality, missingness, and heterogeneity.

Bayesian Statistics

Bayesian methods have become popular for their flexibility to incorporate prior knowledge into statistical models. Bayesian approaches are particularly useful in cancer trials for adaptive designs, where probabilities of one or another decision can be updated while the new data emerge. Adaptability in this sense places Bayesian statistics among the necessary ones in modern clinical research.

7.4. Collaborative Efforts and Cross-Disciplinary Integration

Cancer trials are complex, regarding statistical analysis, and hence call for interdisciplinary collaboration. Statisticians, oncologists, bioinformaticians, and regulatory experts collaborate in the design of robust studies, assurance of quality in data, and interpretation of results with accuracy.

Statisticians

Statisticians are key players in developing innovative trial designs, selecting appropriate analytical methods, and establishing the reproducibility of findings. Their expertise is crucial to overcome several challenges: small sample sizes, high-dimensional data, or nonstandard endpoints.

Interdisciplinary Synergy

The integration of statistical analysis with bioinformatics and computational biology has further enhanced its capability. For example, bioinformaticians work out algorithms for genomic data analysis, and computational biologists provide insight into the biological mechanisms behind the development of statistical models.

Collaboration on Regulations

The FDA and EMA increasingly call for statistical rigor in cancer trials. Various collaborations among researchers and regulators ensure that statistical methods meet evolving standards and support the translation of research findings into clinical practice.

7.5. Future Directions and Potential Innovations

The future of statistical analysis in cancer trials will be transformed with hybrid models, federated learning, and block chain technology. Hybrid models blend the depth of traditional statistics with modern machine learning to bring to life complex relationships within data; this enables precision in patient stratification, dynamic dose optimization, and real-time risk assessment [44]. Federated learning enables institutions to collaboratively analyze data without the sharing of raw datasets, enabling privacy and global collaborations. In return, block chain guarantees data integrity through transparency and secure management of patient consent by recording everything in an immutable manner. Together, these innovations make cancer trials more rigorous, more efficient, and of high ethical standards, thus making treatments faster, more personalized, and more effective.

Hybrid Statistical Models

A marriage of traditional statistical approaches with those that are AI-driven holds exciting potential for improved accuracy and generalizability. For instance, hybrid models that incorporate logistic regression into neural networks can enhance the robustness of predictions while retaining interpretability.

Data-Sharing Ecosystems

Collaborative data-sharing platforms, including cloud-based biobanks, will allow researchers to access a variety of data sources while ensuring patient privacy is maintained. Support for large-scale meta-analyses and generalizable insights will also be facilitated.

Ethics and Social Issues

With the growing intricacy of statistical methodologies, it is vital to ensure that these remain usable and ethical. The transparency of reporting, patient-centric designs, and continuous training of researchers are pivotal in cancer studies for ensuring trust and accountability [45]. Statistical analysis is more than just a technical discipline: It is a linchpin of modern cancer research that bridges the gap between data and actionable insights. By addressing current challenges, embracing technological advancements, and fostering interdisciplinary collaboration, the field can continue to drive breakthroughs that improve patient outcomes and advance the fight against cancer.

8. Conclusion

Statistical analysis in cancer research clinical trials forms the cornerstone for carving the future of oncology. It helps the researcher make sense of complex data sets, assess the efficacy of treatments, and identify patterns driving personalized medicine [49]. Advanced methodologies integrate adaptive trial designs, Bayesian frameworks, and AI-driven analytics to shape the research landscape toward streamlined, more accurate, and truly patient-centered outcomes [50].

Perhaps the most profound impact of the advances in statistics has been the movement toward precision oncology. Techniques such as predictive modeling and subgroup analysis now make it possible to tailor treatments to the profiles of individual patients, thereby optimizing therapeutic outcomes and minimizing toxicities. At the same time, newer tools and big data have extended the dimensions of analysis into new realms of insight regarding patient behavior, disease course, and response to therapy.

However, this field is not without challenges. Issues such as data heterogeneity, high dropouts, and the complexities of new advanced methodologies necessitate further research. The balance of ethical considerations in trial designs must be an innovation that puts the welfare of patients first and observes the strict scientific standards of research. Limitations call for collaborations among statisticians, oncologists, regulatory bodies, and technology experts.

The seamless integration of state-of-the-art technologies into established statistical practices is the future of cancer clinical trials. The field can further enhance its impact on cancer research by embracing hybrid models, fostering data-sharing ecosystems, and continually validating emerging techniques. Ultimately, statistical analysis will continue to be at the leading edge in driving oncology forward with discoveries that give hope to millions of patients worldwide.

Compliance with ethical standards

Acknowledgments

The author, Haoxuan Sun, expresses his heartfelt gratitude to the Department of Mathematics and Statistics at Georgetown University for providing valuable guidance and resources during his academic journey. Special thanks are extended to mentors and colleagues for their unwavering support and insightful discussions that significantly contributed to the completion of this research work.

Disclosure of conflict of interest

The author, Haoxuan Sun, declares no conflict of interest regarding the publication of this research work. All efforts were made to ensure impartiality and the accuracy of the findings presented in this study.

References

[1] Sung, Hyuna, et al. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: a cancer journal for clinicians 71.3 (2021): 209-249.

[2] Xu, Wei, et al. "Statistical fundamentals on cancer research for clinicians: Working with your statisticians." Clinical and Translational Radiation Oncology 27 (2021): 75-84.

[3] Tang, Monica, et al. "Harnessing real-world evidence to advance cancer research." Current Oncology 30.2 (2023): 1844-1859.

[4] Passi, Indu, Sushant Salwan, and Bhupinder Kumar. "US-FDA approved drugs in 2020 and 2021: a review." Mini reviews in medicinal chemistry 23.12 (2023): 1273-1297.

[5] Gupta, Arjun, et al. "Evaluating the time toxicity of cancer treatment in the CCTG CO. 17 trial." JCO Oncology Practice 19.6 (2023): e859-e866.

[6] Akiya, Ippei, Takuma Ishihara, and Keiichi Yamamoto. "Comparison of Synthetic Data Generation Techniques for Control Group Survival Data in Oncology Clinical Trials: Simulation Study." JMIR medical informatics 12.1 (2024): e55118.

[7] Feng, Xiaoshuang, et al. "Cancer stage compared with mortality as endpoints in randomized clinical trials of cancer screening: a systematic review and meta-analysis." JAMA (2024).

[8] Berry, Wendell. The art of the commonplace: The agrarian essays of Wendell Berry. Catapult, 2018.

[9] Spiegelhalter, David J. "Incorporating Bayesian ideas into health-care evaluation." (2004): 156-174.

[10] Topol, Eric J. "High-performance medicine: the convergence of human and artificial intelligence." Nature medicine 25.1 (2019): 44-56.

[11] Duan, L., et al. (2023). "Advanced Statistical Approaches in Cancer Clinical Trials: Machine Learning and Beyond." Journal of Clinical Oncology, 41(7), 1235-1245.

[12] Rubin, Donald B. "The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm." Journal of the american statistical association 82.398 (1987): 543-546.

[13] Sullivan, A., et al. (2023). "Data Integrity and Statistical Reproducibility in Cancer Trials: Challenges and Solutions." Clinical Trials Journal, 31(4), 134-145.

[14] Austin, Peter C. "An introduction to propensity score methods for reducing the effects of confounding in observational studies." Multivariate behavioral research 46.3 (2011): 399-424.

[15] Smith, J., et al. (2023). "Ethical Considerations in Statistical Modeling for Cancer Research." The Lancet Oncology, 24(3), 456-463.

[16] Green, T., et al. (2023). "Statistical Models and AI in Cancer Trials: Addressing the Reliability Gap." Cancer Research and Therapy Journal, 15(2), 98-110. .

[17] GlobalData Report (2023). "Global Trends in Oncology Clinical Trials: Focus on Statistical Analysis." Clinical Trials Arena.

[18] Gupta, Arjun, et al. "Evaluating the time toxicity of cancer treatment in the CCTG CO. 17 trial." JCO Oncology Practice 19.6 (2023): e859-e866.

[19] Le, Toan, et al. "A narrative review of the relationship between head and neck cancer and obstructive sleep apnea: Clinical studies and statistical analysis." Sleep Medicine Reviews (2024): 102010.

[20] Stalpers, L. J. A., and Edward L. Kaplan. "Edward L. Kaplan and the Kaplan-Meier Survival Curve." BSHM Bulletin: Journal of the British Society for the History of Mathematics, vol. 33, no. 2, 2018, pp. 109–135. https://doi.org/10.1080/17498430.2018.1450055.

[21] Cox, David R. "Regression models and life-tables." Journal of the Royal Statistical Society: Series B (Methodological) 34.2 (1972): 187-202.

[22] Ilie, Gabriela, et al. "A comprehensive 6-mo prostate cancer patient empowerment program decreases psychological distress among men undergoing curative prostate cancer treatment: a randomized clinical trial." European urology 83.6 (2023): 561-570.

[23] Kuznia, Sabine, et al. "Efficacy of vitamin D3 supplementation on cancer mortality: Systematic review and individual patient data meta-analysis of randomized controlled trials." Aging Research Reviews 87 (2023): 101923.

[24] Unger, Joseph M., et al. "Persistent disparity: socioeconomic deprivation and cancer outcomes in patients treated in clinical trials." Journal of Clinical Oncology 39.12 (2021): 1339-1348.

[25] Wells, J. Connor, et al. "An analysis of contemporary oncology randomized clinical trials from low/middle-income vs high-income countries." JAMA oncology 7.3 (2021): 379-385.

[26] Rubin, Donald B. "An overview of multiple imputation." Proceedings of the survey research methods section of the American statistical association. Vol. 79. 1988.

[27] Unger, Joseph M., et al. "When offered to participate": a systematic review and meta-analysis of patient agreement to participate in cancer clinical trials." JNCI: Journal of the National Cancer Institute 113.3 (2021): 244-257.

[28] Atkins, Johnique T., et al. "Pre-clinical animal models are poor predictors of human toxicities in phase 1 oncology clinical trials." British journal of cancer 123.10 (2020): 1496-1501.

[29] Simon, Richard. "Roadmap for developing and validating therapeutically relevant genomic classifiers." Journal of Clinical Oncology 23.29 (2005): 7332-7341.

[30] Goodwin, Pamela J., et al. "Effect of metformin vs placebo on invasive disease–free survival in patients with breast cancer: the MA. 32 randomized clinical trials." Jama 327.20 (2022): 1963-1973.

[31] Khodabakhshi, Adeleh, et al. "Feasibility, safety, and beneficial effects of MCT-based ketogenic diet for breast cancer treatment: a randomized controlled trial study." Nutrition and cancer 72.4 (2020): 627-634.

[32] Li, Yimei, et al. "Statistical considerations for analyses of time-to-event endpoints in oncology clinical trials: illustrations with CAR-T immunotherapy studies." Clinical Cancer Research 28.18 (2022): 3940-3949.

[33] Ito, Chiyo, et al. "Misleading reporting (spin) in noninferiority randomized clinical trials in oncology with statistically not significant results: a systematic review." JAMA Network open 4.12 (2021): e2135765-e2135765.

[34] Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." nature 542.7639 (2017): 115-118.

[35] Homer, Victoria, et al. "Early phase clinical trials extension to guidelines for the content of statistical analysis plans." bmj 376 (2022).

[36] Fujimura, Masaki, et al. "Up-regulation of ICH-1 L protein by thromboxane A 2 antagonists enhances cisplatin-induced apoptosis in non-small-cell lung-cancer cell lines." Journal of cancer research and clinical oncology 125 (1999): 389-394.

[37] Leo, Chandra P., Cornelia Leo, and Thomas D. Szucs. "Breast cancer drug approvals by the US FDA from 1949 to 2018." Nat. Rev. Drug Discov 19.11 (2020).

[38] Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. Nature Biomedical Engineering, 2(10), 719-731.

[39] Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2020). The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 2(9), e425-e427.

[40] Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. New England Journal of Medicine, 378(11), 981-983.

[41] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.

[42] Shabani, M., & Marelli, L. (2019). Re-identifiability of genomic data and the implications for data sharing. Nature Reviews Genetics, 20(9), 497-498

[43] Kaissis, G., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2(6), 305-311.

[44] Esteva, A., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24-29.

[45] Bland, J. M., & Altman, D. G. (1998). Survival probabilities (the Kaplan-Meier method). BMJ, 317(7172), 1572-1580.

[46] Liu, Y., et al. (2021). A framework for AI-based clinical trial design in oncology. Nature Cancer, 2(2), 120-128.

[47] Johnson, K. W., et al. (2021). Global adoption of AI in health care: the case for ethical practices. Nature Medicine, 27(1), 23-26.

[48] Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. Epidemiology, 19(5), 640-648.

[49] Schwimmer, C. M., et al. (2020). The role of statistical methods in advancing personalized cancer treatment. Journal of Clinical Oncology, 38(7), 723-731.

Authors short Biography

Haoxuan Sun completed his Master's degree in Mathematics and Statistics at Georgetown University, United States. His academic journey includes expertise in statistical methods and data analysis, with a strong focus on clinical trial data analysis. Haoxuan is dedicated to contributing meaningful insights in applied mathematics and statistics through his research efforts.