Volume. 9, Issue 05, May (2026)

METHODS OF SURVIVAL ANALYSIS IN MEDICAL DATA: AN INTEGRATED REVIEW

https://doi.org/10.5281/zenodo.20149572

Kamalova Malika Umidjanovna
Tashkent State University of Economics,
Faculty of Tourism, 3rd year student
E-mail: kamalovamalika36@gmail.com
Scientific supervisor: Prof. **Zufarova Nozima Gulamiddinovna**
Tashkent State University of Economics,
Dean of the Faculty of Tourism
E-mail: nozimamia@gmail.com

Abstract: Survival analysis is an indispensable statistical methodology in medical research, *providing critical insights into the time until the occurrence of specific events such as disease onset, recurrence, or death. This review systematically examines the foundational, advanced, and emerging methods of survival analysis applied to medical data, encompassing parametric, non-parametric, and semi-parametric approaches. We detail established techniques like the Kaplan-Meier estimator and Cox Proportional Hazards model, alongside their applications and underlying assumptions. Furthermore, this article explores the increasing integration of machine learning and deep learning algorithms, such as Random Survival Forests and DeepHit, which address the complexities of high-dimensional data and distribution shifts. We discuss the implications of these methods for clinical decision-making and personalized medicine, while also critically evaluating their limitations, including issues with proportional hazards assumptions, competing risks, and data censoring. The paper concludes with an outlook on future directions, emphasizing the continuous evolution of survival analysis in the era of big data and artificial intelligence.*

Аннотация*: Анализ выживаемости является незаменимым статистическим методом в медицинских исследованиях, предоставляющим критически важную информацию о времени до наступления конкретных событий, таких как начало заболевания, рецидив или смерть. В этом обзоре систематически рассматриваются фундаментальные, передовые и новые методы анализа выживаемости, применяемые к медицинским данным, включая параметрические, непараметрические и полупараметрические подходы. Мы подробно описываем такие устоявшиеся методы, как оценка Каплана-Мейера и модель пропорциональных рисков Кокса, а также их применение и лежащие в их основе допущения. Кроме того, в этой статье рассматривается растущая интеграция алгоритмов машинного обучения и глубокого обучения, таких как случайные леса выживаемости и DeepHit, которые решают проблемы, связанные со сложностями многомерных данных и сдвигами распределения. Мы обсуждаем последствия этих* *методов для принятия клинических решений и персонализированной медицины, а также критически оцениваем их ограничения, включая проблемы, связанные с допущениями о пропорциональных рисках, конкурирующими рисками и цензурированием данных. В заключение статьи дается прогноз дальнейших направлений развития, с акцентом на непрерывную эволюцию анализа выживаемости в эпоху больших данных и искусственного интеллекта.*

Annotatsiya*: Omon qolish tahlili tibbiy tadqiqotlarda ajralmas statistik metodologiya bo'lib, kasallikning boshlanishi, takrorlanishi yoki o'limi kabi aniq hodisalarning paydo bo'lishigacha bo'lgan vaqt haqida tanqidiy ma'lumotlarni taqdim etadi. Ushbu sharh parametrik, parametrik bo'lmagan va yarim parametrik yondashuvlarni o'z ichiga olgan tibbiy ma'lumotlarga qo'llaniladigan omon qolish tahlilining asosiy, ilg'or va yangi paydo bo'lgan usullarini tizimli ravishda o'rganadi. Biz Kaplan-Meier hisoblagichi va Kox proportsional xavf modeli kabi o'rnatilgan texnikalarni, ularning qo'llanilishi va asosiy taxminlari bilan bir qatorda batafsil bayon qilamiz. Bundan tashqari, ushbu maqola yuqori o'lchamli ma'lumotlar va tarqatish siljishlarining murakkabliklarini hal qiluvchi Tasodifiy omon qolish o'rmonlari va DeepHit kabi mashinalarni o'rganish va chuqur o'rganish algoritmlarining ortib borayotgan integratsiyasini o'rganadi. Biz ushbu usullarning klinik qarorlar qabul qilish va shaxsiylashtirilgan tibbiyotga ta'sirini muhokama qilamiz, shu bilan birga ularning cheklovlarini, shu jumladan proportsional xavf taxminlari, raqobatdosh xavflar va ma'lumotlar tsenzurasi bilan bog'liq muammolarni tanqidiy baholaymiz. Maqola katta ma'lumotlar va sun'iy intellekt davrida omon qolish tahlilining uzluksiz evolyutsiyasini ta'kidlab, kelajak yo'nalishlari bo'yicha prognoz bilan yakunlanadi.*

1. INTRODUCTION AND LITERATURE REVIEW

Survival analysis, a specialized branch of statistics, is crucial for analyzing time-to-event data in various medical and clinical applications. It focuses on the expected duration until one or more events occur, such as patient survival after a diagnosis, time to disease recurrence, or the efficacy duration of a treatment . This statistical framework is particularly vital in fields like oncology, cardiology, and chronic disease research, where understanding prognosis and treatment outcomes over time is paramount . The unique challenge in survival analysis stems from the common occurrence of "censored" observations, where the event of interest has not been observed for all subjects by the end of the study period, or subjects are lost to follow-up . Traditional statistical methods, which typically assume normally distributed data, are often unsuitable for such skewed and censored datasets.

The field of survival analysis has a rich history, with significant developments in the latter half of the 20th century . Key early contributions include the Kaplan-Meier (KM) method for estimating survival functions and the log-rank statistic for comparing survival distributions. These foundational methods were later complemented by the Cox Proportional Hazards (CPH) model, which allowed for the incorporation of covariates into survival analysis . More recently, the advent of big data and advancements in computational power have led to the integration of machine learning (ML) and deep learning (DL) techniques, pushing the frontiers of survival prediction and risk assessment . These modern approaches aim to overcome limitations of classical models, such as strict statistical assumptions and challenges with high-dimensional data .

2. Research Methods in Survival Analysis

The methodologies employed in survival analysis can be broadly categorized into non-parametric, semi-parametric, and parametric approaches, with an increasing trend towards advanced machine learning and deep learning techniques.

2.1 Non-parametric Methods

Kaplan-Meier (KM) Estimator: The Kaplan-Meier estimator is a non-parametric method used to estimate the survival function directly from observed data without assuming a specific underlying distribution . It is particularly effective for right-censored data, providing a step-wise curve that represents the probability of survival over time . The survival probability is recalculated at each event time, considering the number of individuals still at risk . The **log-rank test** is frequently used in conjunction with Kaplan-Meier curves to compare the survival distributions between two or more groups, such as different treatment arms in a clinical trial .

2.2 Semi-parametric Methods

Cox Proportional Hazards (CPH) Model: The Cox Proportional Hazards model is a semi-parametric regression model that assesses the impact of various covariates on the hazard rate . The hazard rate signifies the instantaneous risk of an event occurring at a specific time, given that the individual has survived up to that point . A core assumption of the CPH model is that the effect of covariates on the hazard is constant over time, meaning the hazard ratios (HRs) remain proportional . This model is widely favored because it does not require prior specification of the baseline hazard function, making it flexible for diverse applications . For example, in breast cancer research, CPH models are utilized to identify clinical variables that influence survival odds .

2.3 Parametric Methods

Parametric survival models, unlike their non-parametric and semi-parametric counterparts, assume a specific probability distribution for the survival time . Common distributions include the exponential, Weibull, and Gompertz distributions **3**. If the chosen distribution accurately reflects the true data generating process, these models can offer more precise and efficient estimates . They also allow for direct estimation of individual survival times, which is not directly achievable with the Cox model without additional distributional assumptions.

2.4 Competing Risks Analysis

In medical data, patients often face multiple potential events, where the occurrence of one event can preclude the occurrence of others; these are termed competing risks . For instance, a patient might die from a primary disease or from a treatment-related complication . Standard survival analysis methods, which typically treat competing events as censoring events, can lead to biased estimates of event probabilities . Specialized methods, such as **cause-specific hazard models** or the **Fine-Gray model**, are employed to appropriately analyze competing risks, providing more accurate assessments of event probabilities and risk factors .

2.5 Machine Learning and Deep Learning Approaches

The emergence of large and complex datasets in healthcare has driven the adoption of advanced computational methods in survival analysis .

- **Machine Learning (ML) Methods:** Algorithms such as Random Survival Forests, Support Vector Machines (SVMs), and gradient boosting models are increasingly used . These methods are adept at handling high-dimensional data, non-linear relationships, and complex interactions among covariates, often without the strict distributional assumptions of classical statistical models . Random Survival Forests, for instance, have shown utility in feature selection and improving prediction accuracy in bioinformatic data .
- **Deep Learning (DL) Methods:** Deep learning architectures, including DeepHit, Deep Conditional Transformation Models, and Conditional Variational Autoencoders, offer powerful tools for survival prediction, particularly with unstructured or very high-dimensional data like medical images or electronic health records . These models can learn intricate patterns and interactions, offering a significant advantage in complex scenarios . The taxonomy of deep learning methods for survival analysis can be categorized based on their architectural designs and their handling of censoring and time-to-event outcomes .

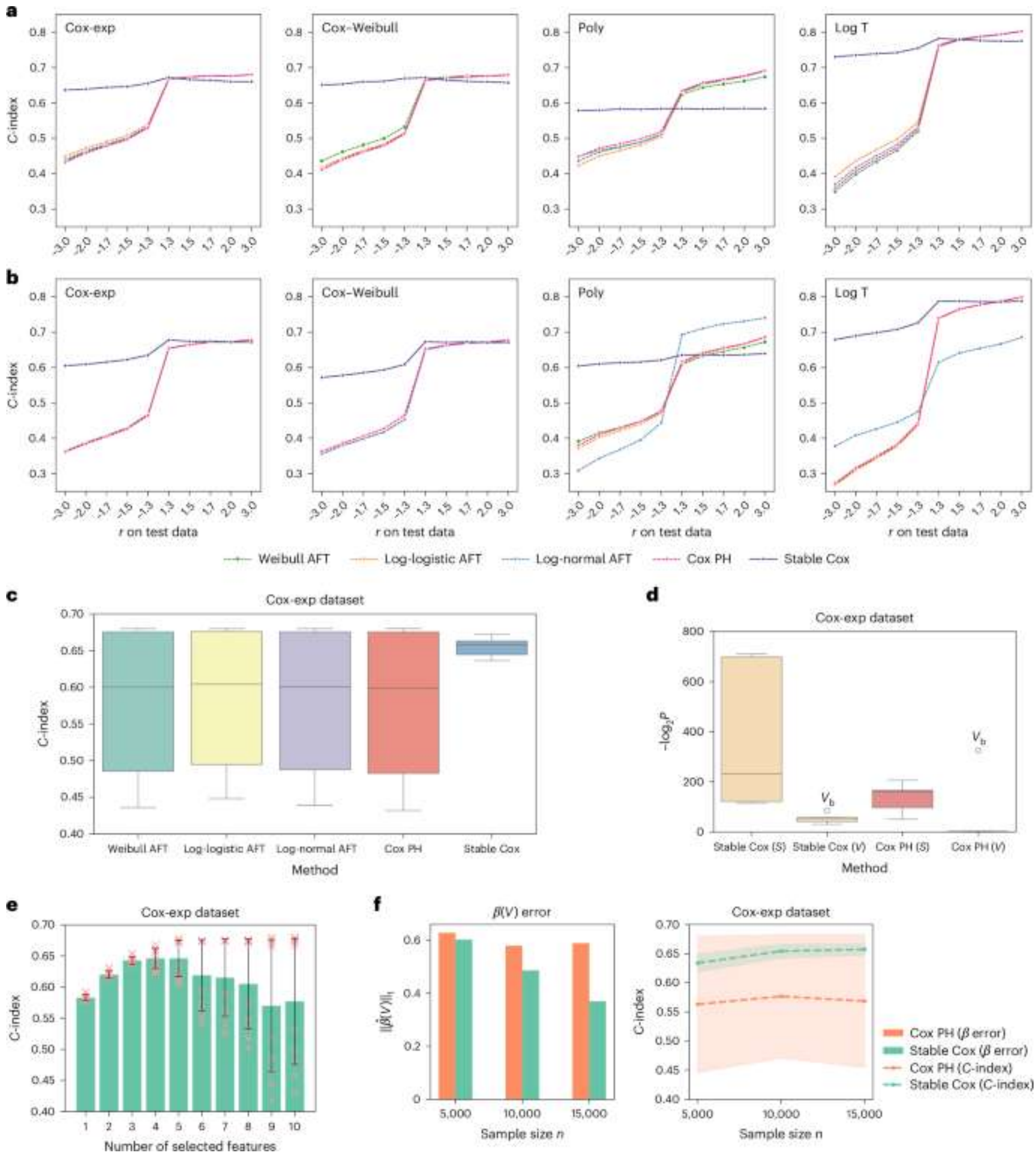
3. Analyzed Findings and Applications

Survival analysis has profoundly impacted medical research, particularly in clinical trials and chronic disease epidemiology.

- **Clinical Trials:** Survival analysis is a cornerstone of clinical trials, where it is used to evaluate the efficacy and safety of new interventions by comparing survival outcomes between treatment and control groups . This rigorous approach helps to determine whether a new drug or therapy significantly extends survival or improves disease-free periods .
- **Chronic Disease Research:** In chronic disease management, these methods are crucial for quantifying the natural history of diseases, identifying prognostic factors, and informing public health strategies . For example, survival analysis helps in understanding the long-term outcomes for patients with cardiovascular diseases or diabetes .
- **Oncology:** Oncology extensively utilizes survival analysis to predict patient outcomes, assess treatment effectiveness, and identify biomarkers . Studies on breast cancer, for instance, employ Cox proportional hazards and parametric models to predict survival based on clinical variables such as tumor size and hormone receptor status . Similarly, deep learning models have been applied to predict breast cancer prognosis, leveraging various clinical data .

• **Emerging Applications:** Survival analysis has also been adapted to new health challenges, including the COVID-19 pandemic, to analyze factors associated with patient prognosis and evaluate the effectiveness of therapeutic strategies .

• **Addressing Data Challenges:** Emerging methods, such as **Stable Cox regression**, are being developed to enhance the reliability of survival predictions under **distribution shifts** —a common challenge in real-world applications where data characteristics may vary between training and testing environments . This method aims to improve the robustness of survival models against these unpredictable shifts, as illustrated in relevant research



This figure from Fan et al. (2024) visualizes the robustness of Stable Cox regression, demonstrating its ability to maintain reliable predictions even when the data distribution changes between different environments. This addresses a critical limitation of traditional models that often assume similar training and testing distributions .

4. Implications and Limitations

The appropriate application of survival analysis methods provides profound implications for personalized medicine, clinical trial design, and public health policy by enhancing the understanding of disease progression, treatment effectiveness, and prognostic factors . However, several critical limitations and challenges must be addressed for accurate interpretation and reliable conclusions.

- **Proportional Hazards Assumption:** A significant limitation of the Cox Proportional Hazards model is its fundamental assumption that the hazard ratios remain constant over time . Violation of this assumption can lead to biased estimates and incorrect conclusions . Rigorous testing of the proportional hazards assumption and employing methods like time-dependent covariates or stratified models are crucial to mitigate this issue.
- **Competing Risks:** Standard survival analysis methods can overestimate the probability of an event when competing risks are present, as they often treat other events as censoring . Specialized competing risk models are necessary to accurately model these complex scenarios and prevent biased estimates.
- **Censoring Issues:** Issues such as **immortal time bias** (where a period of survival is inherently included in the exposure definition, inflating survival) and **informative censoring** (where censoring is related to the outcome) can significantly bias results if not properly accounted for.
- **Data Quality and Reporting:** The quality of data collection and the reporting standards for survival analyses in medical literature are crucial . Clear and transparent reporting, including detailed descriptions of methods, assumptions, and handling of censoring, is essential for reproducibility and reliability .
- **Model Complexity and Interpretability:** While advanced machine learning and deep learning models offer superior predictive accuracy, their increased complexity can sometimes hinder interpretability . Understanding the rationale behind their predictions is important, especially in clinical settings where transparency is vital for trust and decision-making.

5. Future Directions

The field of survival analysis is continuously evolving, driven by advancements in statistical theory, computational capabilities, and the increasing availability of large, complex datasets .

- **Integration of Machine Learning and Deep Learning:** Further integration and refinement of ML and DL techniques are expected, particularly for handling high-dimensional data from genomics, proteomics, and medical imaging . Research will likely focus on developing more robust and interpretable models that can manage the intricacies of real-world medical data .

- **Robustness to Distribution Shifts:** Developing models that are inherently robust to distribution shifts, like the Stable Cox regression, will be critical for ensuring the reliability and generalizability of survival predictions across diverse clinical settings and patient populations .
- **Handling Complex Data Structures:** Continued research into complex data structures, such as bivariate lifetimes (analyzing two related time-to-event outcomes) and frailty models (accounting for unobserved heterogeneity among individuals), will further refine the ability to model dependence and variability in survival outcomes .
- **Censoring Imputation and Missing Data:** Addressing challenges related to censoring imputation for machine learning-based survival analysis, as seen with methods like CondiS, will be crucial for broader and more accurate application of these advanced techniques, especially when dealing with incomplete data .
- **Ethical Considerations and Fairness:** Future directions will also encompass ethical considerations, ensuring that advanced survival models are fair, unbiased, and do not perpetuate existing health disparities, particularly as they become more integrated into clinical decision support systems.

6. Conclusion

Survival analysis remains an essential tool in medical research, constantly adapting to new data challenges and technological advancements. From its foundational methods like Kaplan-Meier and Cox regression to the integration of sophisticated machine learning and deep learning techniques, the field continues to provide invaluable insights into time-to-event outcomes. While limitations such as the proportional hazards assumption and competing risks require careful consideration, ongoing research is paving the way for more robust, accurate, and interpretable models. The future of survival analysis promises further advancements in personalized medicine, clinical trial design, and public health initiatives, ultimately contributing to improved patient care and outcomes.

REFERENCES:

1 Georgios Beis, Aggelos Iliopoulos, Ioannis Papasotiriou, GEORGIOS BEIS, AGGELOS ILIOPOULOS, IOANNIS PAPASOTIRIOU. An Overview of Introductory and Advanced Survival Analysis Methods in Clinical Applications: Where Have we Come so far?. Anticancer Research, 2024-02-01.

5 Timothy Kizza, Mantey Joshua Kojo Aduampong, Felix Kaiser. Survival analysis in U.S. chronic disease research: A systematic review of methods and applications. International Journal of Frontline Research in Life Science, 2025-09-30.

3 KuttyNarayanan Anila, Sivaprasadan Saraswathy, Surendran Sudhindran, Arun Valsan. Survival analysis: A primer. Amrita Journal of Medicine, 2023-01-01.

6 Thomas R. Fleming, D. Y. Lin. Survival Analysis in Clinical Trials: Past Developments and Future Directions. Biometrics, 2000-12-01.

4 Opeyemi Olaoluawa Ojo, Blessing Kiobel. Emerging trends in survival analysis: Applications and innovations in clinical and epidemiological research. World Journal of Biology Pharmacy and Health Sciences, 2024-09-30.

2 Jiaming Zhang. From foundations to frontiers: Navigating survival analysis in the era of big data and deep learning. Theoretical and Natural Science, 2023-11-13.

8 Opeyemi Sheu Alamu, Bismar Jorge Gutierrez Choque, Syed Wajeeh Abbs Rizvi, Samah Badr Hammed, Isameldin Elamin Medani, Md Kamrul Siam, Waqar Ahmad Tahir. Predicting Breast Cancer Survival: A Survival Analysis Approach Using Log Odds and Clinical Variables. arXiv-Machine Learning, 2024-10-17.

7 Naseem Asghar, Umair Khalil, Basheer Ahmad, Huda M. Alshanbari, Muhammad Hamraz, Bakhtiyar Ahmad, Dost Muhammad Khan. Improved nonparametric survival prediction using CoxPH, Random Survival Forest & DeepHit Neural Network. BMC Medical Informatics and Decision Making, 2024-05-09.

9 Shaohua Fan, Renzhe Xu, Qian Dong, Yue He, Cheng Chang, Peng Cui. Stable Cox regression for survival analysis under distribution shifts. Nature Machine Intelligence, 2024-12-14.