# Outcomes and Recurrence in Brain Tumor Patients

Carlos Cantu

# Outcomes and Recurrence in Brain Tumor Patients

This study explores brain tumor progression's complex dynamics, crucial for healthcare advancements, treatment strategies, and patient outcomes. Through stage-based analysis and recurrence pattern investigation, the aim is to decode treatment efficacy and empower personalized care. Brain tumors present challenges due to their diversity and unpredictability. Managing them effectively demands understanding tumor types, treatments, and patient demographics. To overcome data limitations and ethical concerns, we propose using synthetic data mimicking real-world scenarios. Our analysis encompasses patient characteristics, treatment outcomes, and recurrence patterns, employing statistical techniques like descriptive and inferential statistics, survival analysis, and data visualization. Through this approach, we aim to uncover critical factors influencing treatment and recurrence in brain tumor progression.

## Initial Discovery

In the initial exploration, it is observed missing values consistently positioned between the columns "Time_to_Recurrence" and "Recurrence_Site," suggesting a relationship between recurrence timing and site. To streamline data processing, we standardized column names by replacing white spaces with underscores and converted "Tumor_Grade" from Roman numerals to standard numeric format.

The dataset comprises 2000 rows and 11 columns, containing both numerical and string data. Numeric columns include four integers and one floating-point entry, while six columns contain string data. 'Patient_ID' serves as a unique identifier without missing values, ranging from 1 to 2000. Patients' ages, recorded in the 'Age' column, average approximately 56.15 years,

ranging from 42 to 72 years. 'Tumor_Grade' spans from 1 to 4, averaging about 2.433, reflecting varied tumor grades.However, 'Time_to_Recurrence (months)' and 'Recurrence_Site' each exhibit notable missingness at 28.1%, with 562 missing values, indicating incomplete recurrence data. Conversely, 'Survival_Time (months)' contains no missing values, with an average survival period of approximately 34.27 months. 'Treatment_Outcome' contains one word per row with no missing entries.Overall, the dataset appears relatively clean, with no significant anomalies detected during the initial examination. It is now primed for further exploratory data analysis and subsequent tasks.

*Data Summary*

| dataframe | Values |
|---|---|
| Number of rows | 2000 |
| Number of columns | 11 |

*Data Types*

| Column Type | Count |
|---|---|
| string | 6 |
| int64 | 4 |
| float64 | 1 |

*number*

| column_name | NA | NA % | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
|---|---|---|---|---|---|---|---|---|---|---|
| Patient_ID | 0 | 0 | 1000 | 577.5 | 1 | 500.8 | 1000 | 1500 | 2000 | |
| Age | 0 | 0 | 56.15 | 6.078 | 42 | 51 | 56 | 61 | 72 | |
| Tumor_Grade | 0 | 0 | 2.433 | 1.261 | 1 | 1 | 2 | 4 | 4 | |
| Time_to_Recurrence_(months) | 562 | 28.1 | 16.1 | 3.128 | 6 | 14 | 16 | 18 | 36 | |
| Survival_Time_(months) | 0 | 0 | 34.27 | 8.606 | 9 | 28 | 36 | 42 | 72 | |

*string*

| column_name | NA | NA % | words per row | total words |
|---|---|---|---|---|
| Gender | 0 | 0 | 1 | 2000 |
| Tumor_Type | 0 | 0 | 1 | 2000 |
| Tumor_Location | 0 | 0 | 2 | 4000 |
| Treatment | 0 | 0 | 2.7 | 5332 |
| Treatment_Outcome | 0 | 0 | 2 | 4000 |
| Recurrence_Site | 562 | 28.1 | 1.4 | 2876 |

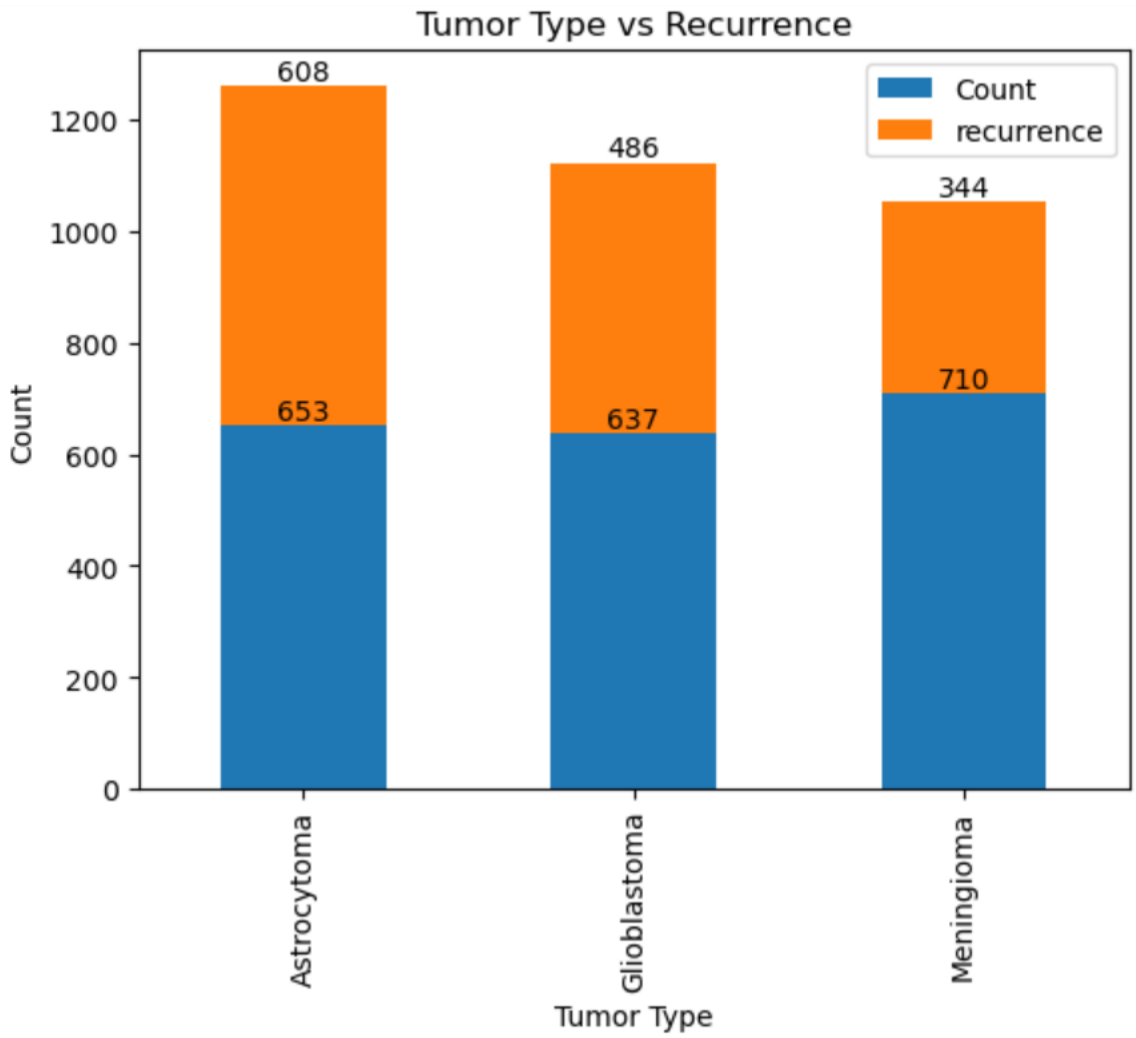
**EDA**

Tumor Type and Recurrence

An essential discovery that emerged was the comprehension of the diverse tumor types present in the dataset and their respective distributions. Among the observed tumor types, ones include Astrocytoma, Glioblastoma, and Meningioma, each representing significant variations in brain tumor pathology and prognosis. For instance, Astrocytoma, derived from astrocytes, and

Glioblastoma, an aggressive form of brain cancer also originating from astrocytes, present distinct challenges in treatment and management. Conversely, Meningioma, arising from the meninges surrounding the brain and spinal cord, offers a unique set of characteristics and considerations.

Interestingly, further investigation revealed that Meningioma was the most represented tumor type within the dataset, with 710 instances recorded. However, despite its prevalence, Meningioma exhibited the fewest tumor recurrences, totaling 344 occurrences. In contrast, both Astrocytoma and Glioblastoma demonstrated similar frequencies, with 653 and 637 instances, respectively. Remarkably, despite their comparable occurrence rates, Astrocytoma displayed the highest number of tumor recurrences, totaling 608 events. These findings underscore the complex interplay between tumor type, occurrence frequency, and recurrence patterns, highlighting the nuanced nature of brain tumor pathology within the dataset.

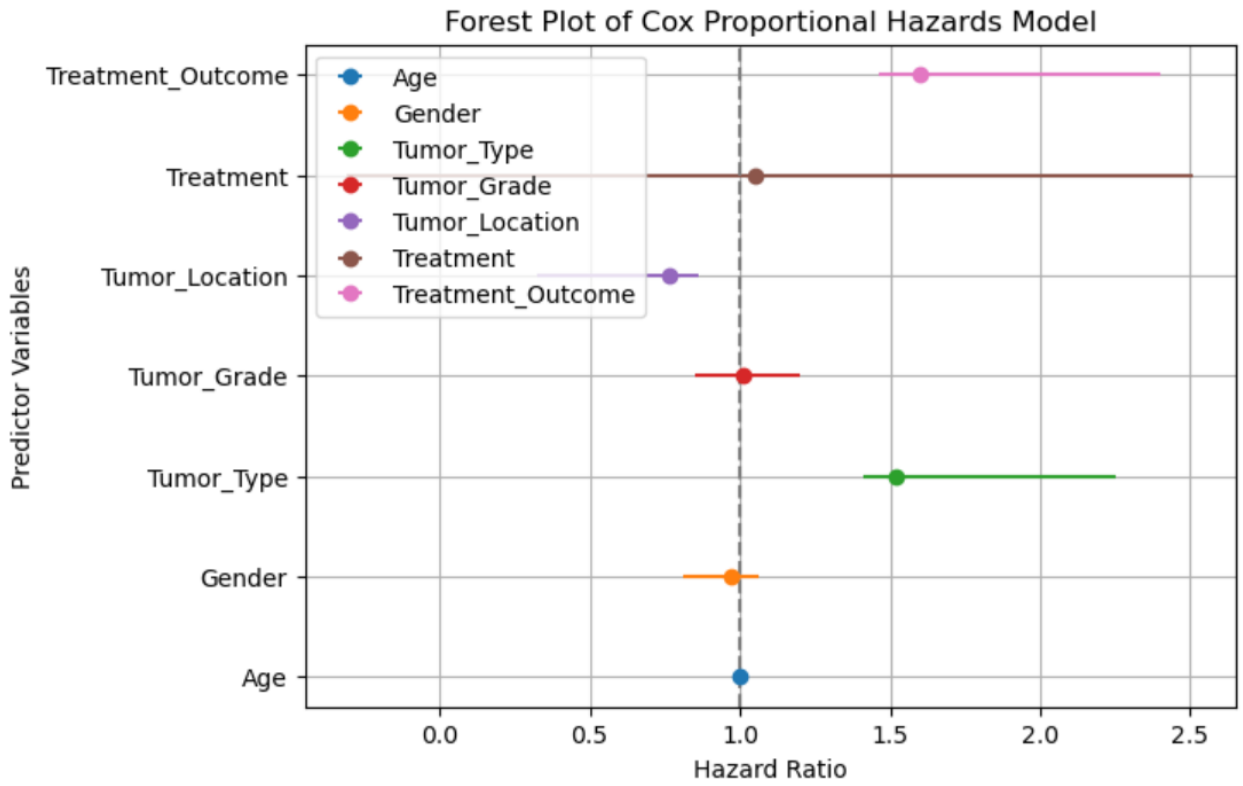



**Model**

Cox Proportional Hazzard Regression

The Cox Proportional Hazards Regression method is used in survival analysis, particularly in oncology, to explore factors affecting time until an event like tumor survival. In investigating brain tumor progression, it reveals several findings. Age and gender show no significant impact on survival time. However, Glioblastoma patients face higher risk (HR = 1.53, $p < 0.005$) of shorter survival. Tumor location, especially in the temporal lobe, predicts better outcomes (HR = 0.43, $p < 0.005$). Various treatments affect survival differently; Surgery alone

(HR = 2.24, p < 0.005) or combined approaches like Surgery with Chemotherapy (HR = 1.87, p < 0.005) have higher hazard ratios. Notably, Surgery with Radiation therapy stands out with a lower hazard ratio (HR = 0.26, p = 0.18), indicating better survival outcomes. Treatment responses significantly influence survival, with Partial response (HR = 0.27, p < 0.005), Progressive disease (HR = 0.43, p < 0.005), or Stable disease (HR = 0.38, p < 0.005) showing lower hazard ratios compared to no response. These factors underscore the complexity of brain tumor management, with tumor type, location, treatment type, and outcome being crucial predictors of survival time. Conversely, age, gender, and tumor grade show no significant effects on survival outcomes.




**Conclusion**

In conclusion, the investigation into the dynamics of brain tumor progression has revealed several critical insights that can significantly impact treatment strategies and patient outcomes. Through a meticulous analysis of a comprehensive dataset, we have elucidated the intricate interplay between various factors such as tumor type, location, treatment modalities, and patient demographics.

Firstly, it has been observed that tumor type plays a pivotal role in both occurrence frequency and recurrence patterns. While Meningioma was the most prevalent tumor type, it exhibited the fewest recurrences, contrasting with Astrocytoma and Glioblastoma, which demonstrated higher recurrence rates despite comparable occurrence frequencies. These findings underscore the complexity of brain tumor pathology and highlight the need for tailored treatment approaches based on tumor type.

Furthermore, the Cox Proportional Hazards Regression analysis unveiled key predictors of survival time. Tumor type emerged as a significant determinant, with patients diagnosed with Glioblastoma facing a notably higher risk of shorter survival times. Additionally, tumor location, treatment modalities, and treatment outcomes significantly influenced survival outcomes. For instance, tumors located in the temporal lobe exhibited improved survival outcomes, while treatment approaches combining surgery with radiation therapy showed the most promising results in prolonging survival.

Conversely, age, gender, and tumor grade did not exhibit significant effects on survival outcomes in our analysis, emphasizing the multifaceted nature of brain tumor progression and the need to focus on tumor-specific characteristics and treatment responses.

Overall, the findings underscore the importance of personalized treatment strategies tailored to individual patients' tumor characteristics and treatment responses. By deciphering the dynamics of brain tumor progression, this study provides actionable insights for clinicians, empowering them to make informed decisions and optimize patient care in the challenging landscape of brain tumor management.

**Assumptions**

In this analysis, it was assumed that all patients in the dataset would eventually succumb to the disease. This assumption provided a framework for using survival time as a relevant metric to understand the dynamics of brain tumor progression. This assumption allows the analysis to use survival time as a relevant metric for understanding of brain tumor progression, as it captures the duration from diagnosis to the inevitable outcome of the disease.

**Limitations and Challenges**

Navigating through challenges and acknowledging limitations is important in conducting meaningful research on brain tumor progression. Limitations include concerns regarding dataset representativeness, particularly its generalizability to broader populations, and the presence of missing values that may compromise the accuracy. Moreover, assumptions of treatment homogeneity overlook potential variations in treatment responses among patient subgroups. Challenges among ensuring data quality, interpreting results from advanced statistical methods accurately. Balancing these factors is essential to ensure the reliability and ethical integrity of research findings in this domain.

**Recommendation and Plan**

Recommendations necessitates a carefully crafted plan. This plan would involve steps such as thorough data preprocessing to address missing values and ensure data compatibility. Advanced statistical analyses, including survival analysis techniques, would identify significant predictors of survival time and recurrence patterns. Additionally, the development of predictive models for recurrence time could offer valuable insights into personalized prognosis and treatment planning for brain tumor patients. Furthermore, patient-centered approaches, including involving patients in decision-making processes, can ensure that treatment plans align with individual preferences and values.

**Ethical Consideration**

In the context of this analysis, where the dataset utilized is synthetic, ethical considerations primarily revolve around the responsible use of data and ensuring the integrity of the research process. While synthetic data mitigates concerns regarding patient privacy and confidentiality, ethical principles still dictate the need for transparency, honesty, and rigorous methodology.

# Citations

The Goan Panda. (2024). Brain Tumor Stage-Based Recurrence Patterns. Retrieved from https://www.kaggle.com/datasets/thegoanpanda/brain-tumor-stage-based-recurrence-patterns

Harrell, Frank E. Jr. "Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis." Springer, 2015.

# Appendix

## Cox Proportional Hazards Regression Analysis

Survival analysis methods offer a versatile framework for assessing multiple risk factors simultaneously, akin to approaches in multiple linear and logistic regression analyses. Among these methods, Cox proportional hazards regression stands out as a popular technique for examining the influence of various risk factors on survival time. In this model, the hazard rate serves as the measure of effect, representing the risk of experiencing the event of interest given survival up to a specific time. While probabilities typically range from 0 to 1, hazards denote the expected number of events per unit of time and can exceed 1 in certain circumstances. The hazard ratio, akin to an odds ratio in logistic regression, facilitates group comparisons regarding their hazards. Estimating the hazard ratio involves assessing the ratio of observed to expected events in two independent comparison groups, often computed through the log-rank test. This abstract underscores the utility of survival analysis in elucidating complex relationships between risk factors and survival outcomes, offering valuable insights for various fields, including healthcare, epidemiology, and social sciences.

$$HR = \frac{\Sigma O_{Exp,t} / \Sigma E_{Exp,t}}{\Sigma O_{Unex,t} / \Sigma E_{Unex,t}} = \frac{\Sigma O_{treated,t} / \Sigma E_{treated,t}}{\Sigma O_{control,t} / \Sigma E_{control,t}}$$

## Questions

1. Can you explain how you ensured the accuracy and reliability of the synthetic dataset used in your analysis?
2. What specific statistical methods did you employ to identify predictors of survival time and recurrence patterns?
3. How did you address missing values in the dataset, and what impact did these missing values have on your analysis?
4. Can you elaborate on the assumptions made in your analysis and how they may have influenced the findings?
5. What were the main limitations and challenges encountered during your research, and how did you mitigate them?
6. Could you provide more details on the predictive models developed for recurrence time and their performance metrics?
7. How did you ensure that patient-centered approaches were incorporated into the analysis and interpretation of the results?
8. What ethical considerations guided your research, particularly concerning the use of synthetic data and patient privacy?
9. Can you discuss any potential biases in the dataset or analysis methods used, and how you addressed them?
10. What are the practical implications of your findings for clinicians, policymakers, and patients in the field of brain tumor management?