Treatment Effectiveness in Patients with Prostate Cancer

Jennifer Cho, Narya Elhawary, Minsol Seo

Due 2025-12-05

Abstract

This project aims to analyze which treatment option for prostate cancer leads to the highest recovery rate among patients. Prostate cancer is one of the most common cancers among men and is ranked as the "second most frequently diagnosed cancer"(Rawla, 2019). It also remains the "fifth leading cause of death worldwide" (Rawla, 2019), making the identification of the most effective treatment essential. As research in cancer therapy continues to expand, it is imperative to compare the overall effectiveness of different treatment methods, including hormonal therapy, radiation therapy, surgery, and combination therapies, to determine whether a single or combined approach yields better recovery outcomes. This study will evaluate tumor size, initial PSA levels, and biopsy Gleason scores over a one-year period to compare patient recovery and cancer recurrence. Statistical analyses will identify patterns of effectiveness among various treatments and determine which method best improves survival rates while reducing recurrence. It is hypothesized that combination treatments will decrease survival and decrease recurrence, whereas single therapies may yield higher recurrence rates but increased survival rates. By analyzing these relationships, this project seeks to highlight how treatment choice influences recovery rates, recurrence, and long-term survival outcomes for prostate cancer patients.

1) Introduction:

The primary purpose of this analysis is to compare and quantify the long-term differential effectiveness of major treatment modalities for prostate cancer—specifically surgery, radiotherapy, hormonal therapy, and a combination of hormonal therapy with radiotherapy. The goal is to apply meticulous statistical methods and rigorous data analysis to determine which specific treatment provides the best outcomes for patient survival and recovery, as well as whether combining therapies is more effective than single-treatment approaches. This study directly addresses the central research question: "What is the differential effectiveness of primary treatment modalities on clinical survival rates and biochemical recurrence outcomes for prostate cancer patients after controlling for prognostic factors?" The analysis is designed to benefit patients by identifying treatments with the most durable effectiveness, while also assisting families and clinicians in making informed decisions based on reliable statistical evidence. By providing outcome-based comparisons, this research aims to guide more personalized treatment strategies tailored to each patient's risk profile and cancer stage. Patient-specific factors, including initial cancer stage (severity and aggressiveness), overall health status, tumor size, initial PSA levels, and biopsy Gleason scores, will be considered. A multivariable regression model will be employed to isolate the effectiveness of treatment type while controlling for these prognostic factors to test the hypothesis that significant differences exist in long-term outcomes among first-line treatments..

2) Data:

The data used for this analysis were retrieved from the Zenodo repository, a well-known open-access repository mainly used by science communities. A dataset we're using, titled "Comprehensive Clinical, Pathological and Follow-up Dataset of Prostate Cancer Patients," was collected and published by Mert Başaranoğlu at

Mersin Üniversitesi Hastanesi (Mersin University Hospital) in southern Turkey. It contains detailed information on factors related to Prostate Cancer, including 30 variables spanning four broader categories: Clinical Parameters, Treatment Information, Pathological Outcomes, and Follow-up Data.

2-1) Dataset Variables:

Table 1: Initial Diagnosis Variables

Variable_Name	English_Translation	Type_of_Data
Hasta_ID	Patient ID	Categorical
Yas	Age	Discrete
Tani_Tarihi	Diagnosis Date	Date
PSA_Tani	Serum Prostate-Specific Antigen (PSA) level at diagnosis (ng/mL)	Continuous
Klinik_Evre	Clinical cT-Stage determined by pre-treatment examinations (cT1c < cT2a < cT2b < cT2c < cT3a < cT3b for increasing extent of tumor invasion)	Ordinal/Categorical
Biyopsi_Gleason	Biopsy Gleason Score (3+3 < 3+4 < 4+3 < 3+5 < 4+4 < 4+5 < 5+4 < 5+5, higher score indicates higher aggressiveness)	Ordinal/Categorical
Risk_Grubu	Risk Group Classification (1 for Low, 2 for Intermediate, 3 forHigh)	Ordinal/Categorical

Table 2: Risk Factors

Variable	In_English	Type_of_Data
Albumin	ASerum albumin level (g/dL). Indicator of nutritional status and systemic health	Continuous
Lenfosit	Lymphocyte (Immune system component) Count	Discrete
CRP	C-Reactive Protein (mg/L). Indicator of inflammation	Continuous
NLR	Neutrophil-to-Lymphocyte Ratio. A prognostic indicator for systemic inflammation and cancer aggressiveness.	Continuous
CALLY_Index	CALLY Index. A composite index, likely related to inflammation or blood components.	Continuous
Komorbidite_Skor	Comorbidity Score indicating the severity of other co-existing chronic diseases (0 (No comorbidities) < ... < 5 (Severe comorbidities))	Ordinal/Categorical

Treatment Information

Table 3: Treatment Information

Variable	In_English	Type_of_Data
Tedavi_Tipi	Main Treatment Type received (1 for Radical Prostatectomy, 2 for Radiotherapy/RT, 3 for Hormone Therapy, 4 for Combination of Radiotherapy and Hormone Therapy)	|Categorical
Tedavi_Tarihi	Treatment Date	Date
RT_Dozu	Total Radiation Dose (in Gy), if radiotherapy was performed	Continuous
ADT_Tipi	Androgen Deprivation Therapy (ADT, hormone therapy) Type used	Categorical
ADT_Suresi	ADT(hormone therapy) Duration	Continuous

Table 4: Pathological Markers

Variable	In_English	Type_of_Data
Patolojik_Evre	Final Tumor Pathological Stage determined after surgery on the removed tissue (pT2a < pT2b < pT2c < pT3a < pT3b < pT4, NaN indicates patient did not undergo surgery)	Ordinal/Categorical
Cerrahi_Sinir	Surgical Margin Status indicating if cancer cells were present at the edge of the removed tissue. Crucial for recurrence prediction (0: Negative, 1: Positive, NaN indicates patient did not undergo surgery).	Binary/Categorical
Final_Gleason	Final Gleason Score confirmed from the final excised tissue (3+3 < 3+4 < 4+3 < 3+5 < 4+4 < 4+5 < 5+4 < 5+5, higher score indicates higher aggressiveness)	Ordinal/Categorical

Table 5: Follow-up & Outcomes

Variable	In_English	Type_of_Data
PSA_Nadir	The lowest PSA level reached after treatment (ng/mL). A lower nadir generally indicates better treatment success	Continuous
PSA_Takip_3ay / 6ay / 12ay	Follow-up PSA levels (ng/mL) measured at at 3/6/12 Months	Continuous
BCR_Durum	Biochemical Recurrence (BCR) Status whether the PSA level rise above a recurrence threshold? (True for Recurrence occurred, False for no Recurrence occurred)	Binary/Categorical

BCR_Tarihi	Date when biochemical recurrence was confirmed	Date
Metastaz_Durum	Metastasis Status whether distant metastasis occur during follow-up? (0 for No, 1 for Yes)	Binary/Categorical
Metastaz_Tarihi	Date when metastasis was confirmed	Date
Son_Durum	Patient's Survival Status at the last follow-up (0 for Alive, 1 for Deceased) |	inary/Categorical |
Son_Takip_Tarihi	Date of the last recorded patient information.	Date

2-2) Data Cleaning

To make data sets more consistent, we then decided to delete unnecessary information by removing the corresponding columns. Afterward, converting all categorical variables into factor type was required to ensure that our program, R(Studio), correctly interprets them as discrete categories rather than as continuous values or even strings. Moreover, this foreign data set has named its critical variables in another language that we need to rename in English. Lastly, addressing missing variables and visualizing were involved in each of them in various forms. These process allow us to observe that missing values were only found for the Hormone Type (ADT_Tipi) variable in patients who did not receive Hormone Therapy (as it is not applicable). Also, information is only present for the specific treatment a patient received (e.g., Radiation Dose is missing for surgery patients and vise versa).

3) Visualization:

We use preliminary visualizations and numerical summaries in order to prepare and interpreting the two multivariate logistic regression models and one linear regression, as this step is essential for diagnosing data quality, identifying confounding factors, and assessing the signal strength before the formal statistical analysis begins. We've established a few bar graphs on the frequency of treatment type, frequency of biochemical recurrence status, survival status, and risk factors. The box plots on PSA level at diagnosis and after treatment will be used to directly confound our numerical data among different treatment groups, as well as interpreting information from risk factors through our numerical summary.

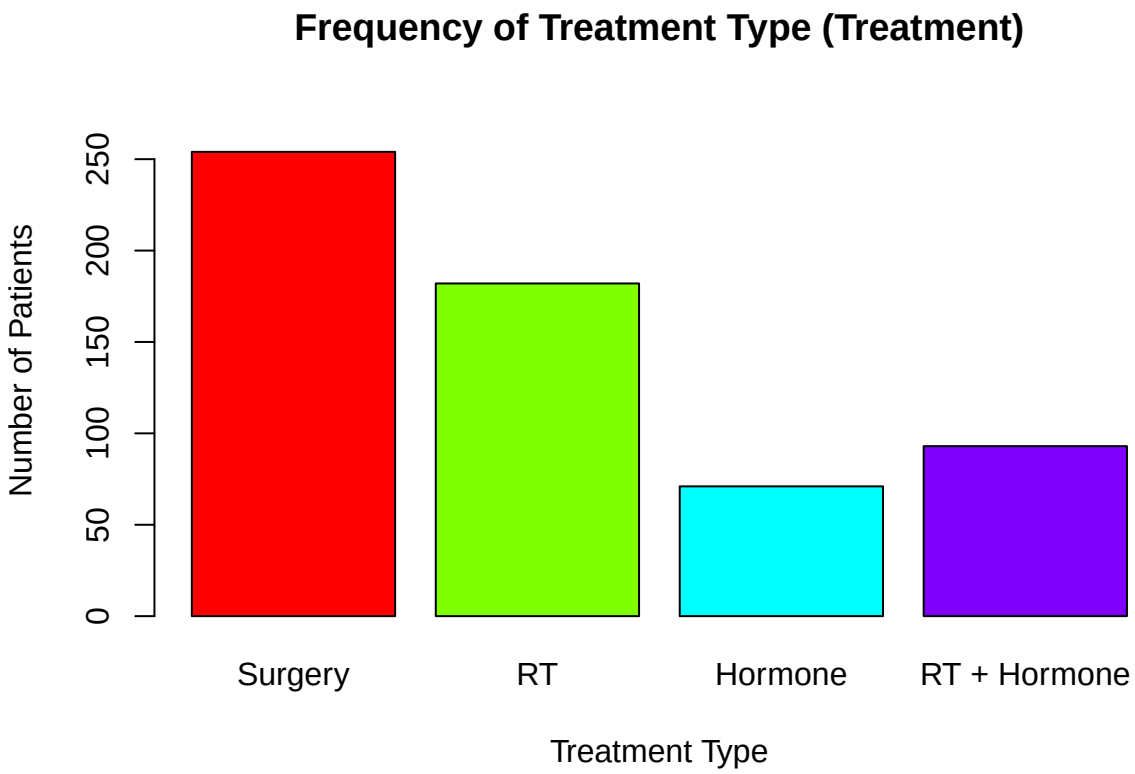


Figure 1: Illustration of distribution of the four main treatment groups across the patient cohort.

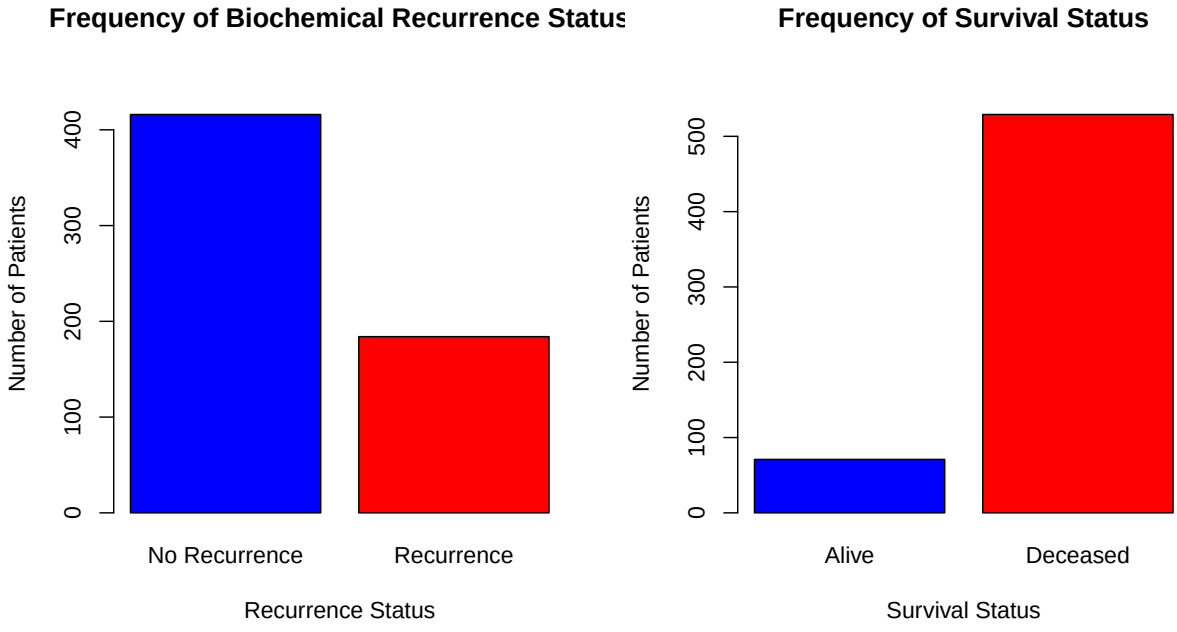


Figure 2: Side-by-side bar plots for Biochemical Recurrence (BCR) and Survival Status to display the main outcomes.

As shown in Figure 2, the data exhibit a low overall recurrence rate of Biochemical Recurrence (BCR), with the majority of patients (69.3%) successfully avoiding recurrence, which is represented as 'False'. However, the surprising and ironic cohort is Survival Status, where 0 = Alive and 1 = Deceased, showing that the survival rate is significantly low, with only 11.8% alive.

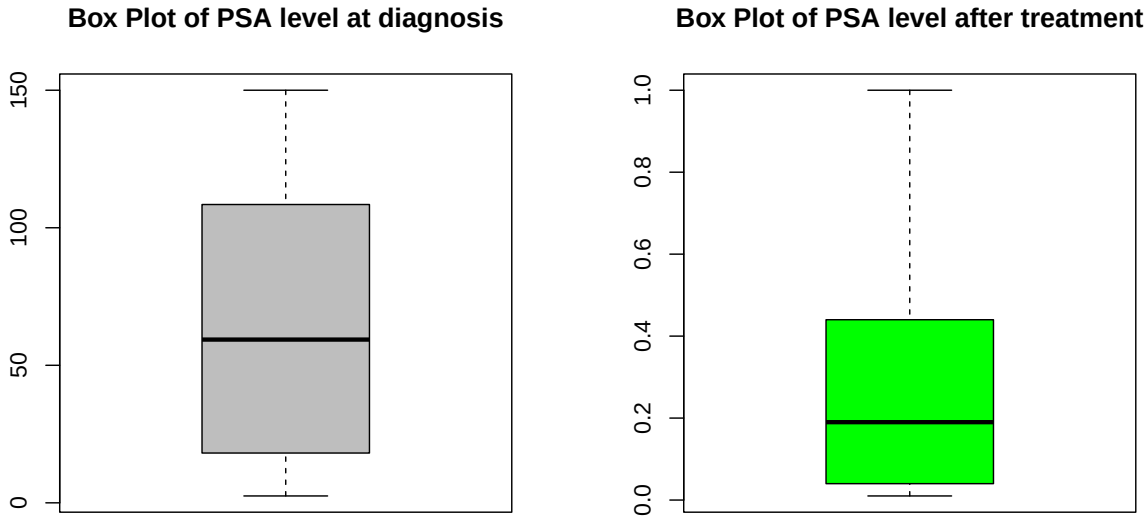


Figure 3: Comparison of PSA level before and after treatment on Diagnosis

Figure 3 displays the distributions and central tendencies of Prostate-Specific Antigen (PSA) levels at the time of diagnosis and after primary treatment. The box plots and numerical summaries illustrate a significant and positive effect of the treatment modalities on reducing PSA levels in the cohort. Before treatment (PSA_before), the median PSA level was 59.35 ng/mL, with values ranging up to 150.00 ng/mL. Following treatment (PSA_after), the median level dropped dramatically to just 0.1900 ng/mL, with most values lying significantly below the pre-treatment range. This substantial decrease in both the median and quartiles (as evidenced by the significant drop from the 1st Quartile of 18.10 to 0.0400 indicates that the treatments were immediately successful in reducing the primary tumor burden and systemic PSA activity in the majority of patients. This observation aligns with clinical expectations, in which successful treatment should drive PSA levels to near-undetectable levels (nadir).

4) Analysis aims:

The main purpose of our analysis is to compare and quantify the long-term discriminatory effects of major first-line treatment methods for patients with prostate cancer. Some of the analyses we aim to complete will include linear regression models, box plot visualizations, R plots, and numerical summaries. We will apply statistical tests such as ANOVA and t-tests to compare treatment effectiveness across different clinical stages, Gleason scores, and recovery outcomes. If prostate cancer is less aggressive (lower Gleason score and early clinical stage), we expect surgery to yield the best outcomes due to its practicality and reduced long-term side effects compared to radiation or hormonal therapy. To test this, we will use box plots to visually compare recovery rates and PSA level reductions among treatment groups, followed by a one-way ANOVA to test for significant differences in mean recovery rates. We hypothesize that patients with localized, low-stage tumors (Stages 1–2) receiving surgery will have significantly higher recovery rates and lower recurrence probabilities. If the cancer is more aggressive or has metastasized, we hypothesize that hormonal therapy will be more effective because hormones circulate through the bloodstream and can target systemic tumor growth. A

multivariate logistic regression model will be used to assess the relationship between treatment type and recurrence or metastasis, controlling for PSA level, tumor size, and patient age. For early-stage cancers (Stages 1–2), radiation therapy may also be effective as a minimally invasive treatment. We will visualize PSA levels and Gleason scores across radiation and surgery groups using box plots, and conduct a t-test to evaluate whether differences in mean PSA reduction between these groups are statistically significant. In advanced stages (Stages 3–4), a combination of hormonal and radiation therapy is expected to provide the best results. We will test this using multivariate logistic regression models for survival and recurrence outcomes, hypothesizing that combination therapy will be associated with the highest survival rates but also with greater treatment intensity and side effects. Overall, these analyses will allow us to statistically evaluate how treatment modality interacts with cancer aggressiveness, Gleason score, and clinical stage to influence recovery, recurrence, and long-term survival outcomes.

4-1) Chi-squared Test on Treatment vs. BCR

Table 6: 4.1 Association between Treatment Type and Biochemical Recurrence (BCR)

	Treatment Type	Recurrence Count (False)	Recurrence Count (True)	No Recurrence (%) (False)	Recurrence (%) (True)
Radical Prostatectomy	Radical Prostatectomy	189	65	74.4	25.6
Radiotherapy (RT)	Radiotherapy (RT)	123	59	67.6	32.4
Hormone Monotherapy	Hormone Monotherapy	40	31	56.3	43.7
Combination (RT + Hormone)	Combination (RT + Hormone)	64	29	68.8	31.2

```
## Chi-squared = 8.9915, df = 3, p-value = 0.0294
```

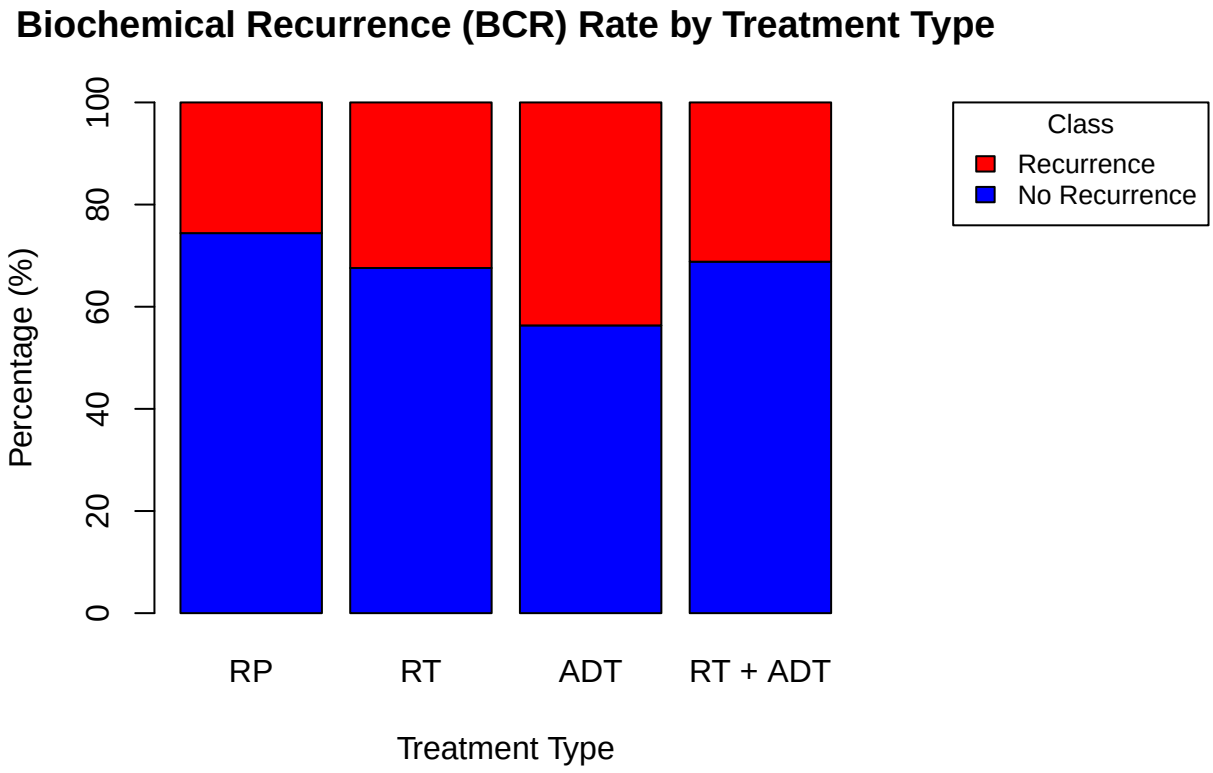
H0: Treatment Type and Biochemical Recurrence (BCR) Status are independent. (The recurrence rate is the same across all treatment groups)

Ha: Treatment Type and Biochemical Recurrence (BCR) Status are not independent. (The recurrence rate is significantly different for at least one treatment group)

Since the p-value (0.0294) is less than the significance level (a=0.05), we reject the null hypothesis (H0). There is a statistically significant association between the type of treatment a patient receives and the likelihood of experiencing Biochemical Recurrence (BCR).

Highest Recurrence Rate: Treatment Type 3 (Hormone Therapy (ADT) Monotherapy) showed the highest biochemical recurrence rate at 43.7%.

Lowest Recurrence Rate: Treatment Type 1 (Radical Prostatectomy) showed the lowest recurrence rate at 25.6%.



4-2) Chi-squared Test on Treatment vs. Survival

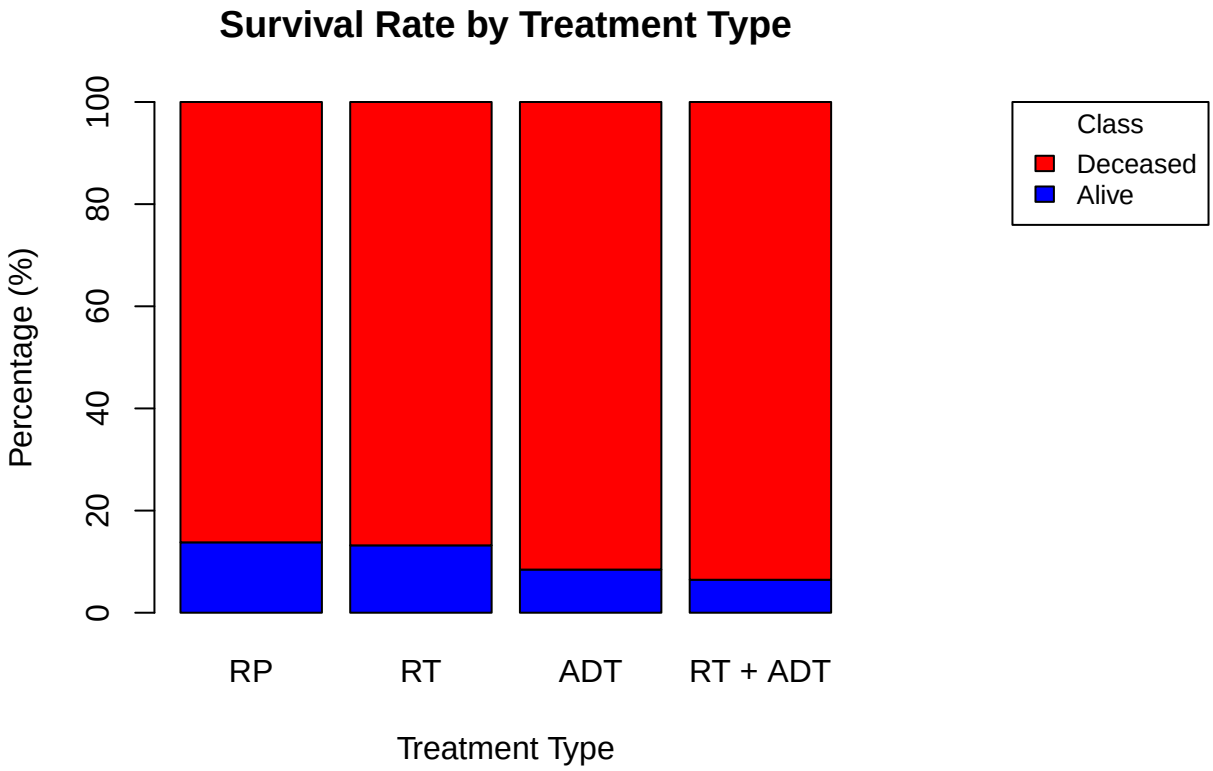
Table 7: Table 4.2: Association between Treatment Type and Survival Status

	Treatment Type	Count (Alive)	Count (Deceased)	Survival (%)	Mortality (%)
Radical Prostatectomy	Radical Prostatectomy	35	219	13.8	86.2
Radiotherapy (RT)	Radiotherapy (RT)	24	158	13.2	86.8
Hormone Monotherapy	Hormone Monotherapy	6	65	8.5	91.5
Combination (RT + Hormone)	Combination (RT + Hormone)	6	87	6.5	93.5

```
## Chi-squared = 4.6021, df = 3, p-value = 0.2034
```

H0: Treatment Type and Survival Status are independent. (The survival rate is the same across all treatment groups) Ha: Treatment Type and Survival Status are not independent. (The survival rate is significantly different for at least one treatment group)

Since the p-value (0.2034) is greater than the significance level (a=0.05), we fail to reject the null hypothesis (H0). There is no statistically significant association between the type of treatment a patient receives and the likelihood of their Survival Status (Alive vs Deceased).



4-3) ANOVA Test, Treatment vs PSA Difference

Table 8: Mean PSA Reduction (Delta PSA) by Treatment Type

Treatment Type	Mean Reduction (PSA_before - PSA_after)
1 (Surgery)	53.905
2 (RT)	68.503
3 (Hormone Therapy)	83.659
4 (RT + Hormone)	70.428

```
## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
## resized.
```

Table 9: Table 4.3: Tukey's HSD Post-Hoc Test for Delta PSA

Group 1	Group 2	Mean Difference	p-value	Significant (p < 0.05)
1 (Surgery)	2 (RT)	14.598	0.0056	Yes
1 (Surgery)	3 (Hormone Therapy)	29.754	0.0000	Yes

1 (Surgery)	4 (RT + Hormone)	16.523	0.0151	Yes
2 (RT)	3 (Hormone Therapy)	15.157	0.0821	No
2 (RT)	4 (RT + Hormone)	1.925	0.9874	No
3 (Hormone Therapy)	4 (RT + Hormone)	-13.232	0.2536	No

H0: The mean change in PSA (Delta PSA) is the same across all four Treatment Type groups

Ha: The mean change in PSA (Delta PSA) is significantly different for at least one treatment group

Since the p-value (2.66e-06) is less than the significance level (a=0.05), we reject the null hypothesis (H0). There is a statistically significant difference in the mean change in PSA (Delta PSA) among the different treatment groups.

The mean PSA change (Delta PSA, where a larger value indicates a greater reduction) in the Radical Prostatectomy (Treatment 1) group is significantly smaller than the mean change in all other treatment groups (2, 3, and 4).

Largest Mean Difference: Treatment 3 (Hormone Therapy Monotherapy) showed the largest positive mean difference (Delta PSA = 29.755) compared to Treatment 1, indicating the greatest average PSA reduction after treatment.

No Significant Difference: There is no significant difference in the mean PSA Difference among the three non-surgical groups (Treatment 2, 3, and 4).

4-4)

5) Conclusion:

Work Cited (Used in Abstract)

Rawla P. (2019). Epidemiology of Prostate Cancer. World journal of oncology, 10(2), 63–89. https://doi.org/10.14740/wjon1191