Predicting Thyroid Cancer Recurrence

Joyce Gill, Matthew Billings-Chiu

2025-05-02

Contents

0.1 Introduction

Cancer is the unregulated proliferation of cells which consume dangerous and disproportionately high amounts of energy from the body. In the United States, cancer is the leading cause of death for people under 65 years old, which calls for more research (Cancer.gov 2025). This study focuses on thyroid cancer, which is projected to account for approximately 44,000 new cases this year. Unlike many other cancer types with significantly declining mortality rates, thyroid cancer has exhibited a slight upward trend in mortality. (Cancer.gov, 2025).

Luckily, thyroid cancer is often treatable, with common interventions including surgical removal of the thyroid and Radioactive Iodine therapy (RAI), which selectively targets thyroid tissues via systemic administration of radioactive iodine (Mayo Clinic 2024). However, treatment response can vary, and understanding the relationship between initial treatment outcome and recurrence is essential for guiding clinical follow-up and secondary interventions.

This study investigates the following research question: Controlling for age, gender, prior radiotherapy, and clinical risk classification, is initial treatment response a statistically significant predictor of thyroid cancer recurrence? We hypothesize that the initial treatment response is a statistically significant predictor of thyroid cancer recurrence, holding other variables we suspect to be significant constant.

Our findings indicate that treatment outcome is a statistically significant predictor of recurrence when controlling for age, gender, radiotherapy history, and risk classification. We also explore the influence and interrelationships of additional explanatory variables to provide a more comprehensive understanding of recurrence risk in differentiated thyroid cancer patients.

0.2 Methods

Our dataset was sourced from Kaggle and originates from a published article by Hamadan University in the European Archives of Oto-Rhino-Laryngology. It contains data on 383 thyroid cancer patients, each of whom was followed for at least 10 years, with records spanning over a 15-year period. The dataset includes information related to patient demographics, treatment history, and clinical outcomes.

To answer our research question, we examined the following variables:

- Recurrence (Binary Categorical Dependent Variable): Whether or not cancer recurred.
- Age (Quantitative Explanatory Variable): Patient's age in years.
- Gender (Binary Categorical Explanatory Variable): Male or female.
- Radiotherapy (Binary Categorical Explanatory Variable): History of prior radiotherapy (yes or no).
- Risk (3 Level Categorical Explanatory Variable): Cancer risk classification (low, medium, high).
- TreatmentOutcome (4 Level Categorical Explanatory Variable): Initial treatment response (excellent, indeterminate, structural incomplete, biochemical incomplete).

For further analysis, we recoded the Recurrence variable as a numeric binary indicator (0 = no recurrence, 1 = recurrence) and collapsed Risk into a binary variable (high vs. not high). We then fit a multiple logistic regression model using recurrence as the response variable and the following as the predictors: treatment outcome, age, gender, radiotherapy history, and binary risk classification.

$$\text{Recurrence} = \beta_0 + \beta_1 \cdot \text{TreatmentOutcome} + \beta_2 \cdot \text{Age} + \beta_3 \cdot \text{Gender} + \beta_4 \cdot \text{Radiotherapy} + \beta_5 \cdot \text{risk\_binary}$$

#! GRAHAM SAYS WE COULD DISCUSS AN EMPIRICAL LOGIT PLOT SPECIFICALLY FOR AGE SINCE IT IS QUANTITATIVE

Since our primary predictors are categorical, the assumption of linearity in the logit is satisfied by design. The assumption of independence was reasonably satisfied, as each patient was observed and followed individually, with no repeated measures or clustering. The assumption of randomness, however, was only partially met.

Although patients were randomly assigned to training and validation sets, they were originally drawn from a single medical center, limiting the generalizability of the results. As such, inferential statistics should be interpreted as valid within this clinical sample, but not necessarily generalizable to the broader thyroid cancer population without further external validation.

0.3 Analysis and Results

We fit the logistic regression model described above to predict thyroid cancer recurrence. The table below summarizes the model coefficients, standard errors, z-values, and p-values for each term in the model.

Table 1: Logistic Regression for Thyroid Cancer Recurrence

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	22.392	3357.120	0.007	0.995
TreatmentOutcomeExcellent	-3.034	1.089	-2.786	0.005
TreatmentOutcomeBiochemical Incomplete	1.663	0.637	2.612	0.009
TreatmentOutcomeStructural Incomplete	5.671	0.875	6.478	0.000
Age	0.030	0.020	1.532	0.125
GenderM	1.043	0.646	1.614	0.107
RadiotherapyYes	-10.634	2414.759	-0.004	0.996
risk_binaryNot High	-26.058	3357.119	-0.008	0.994

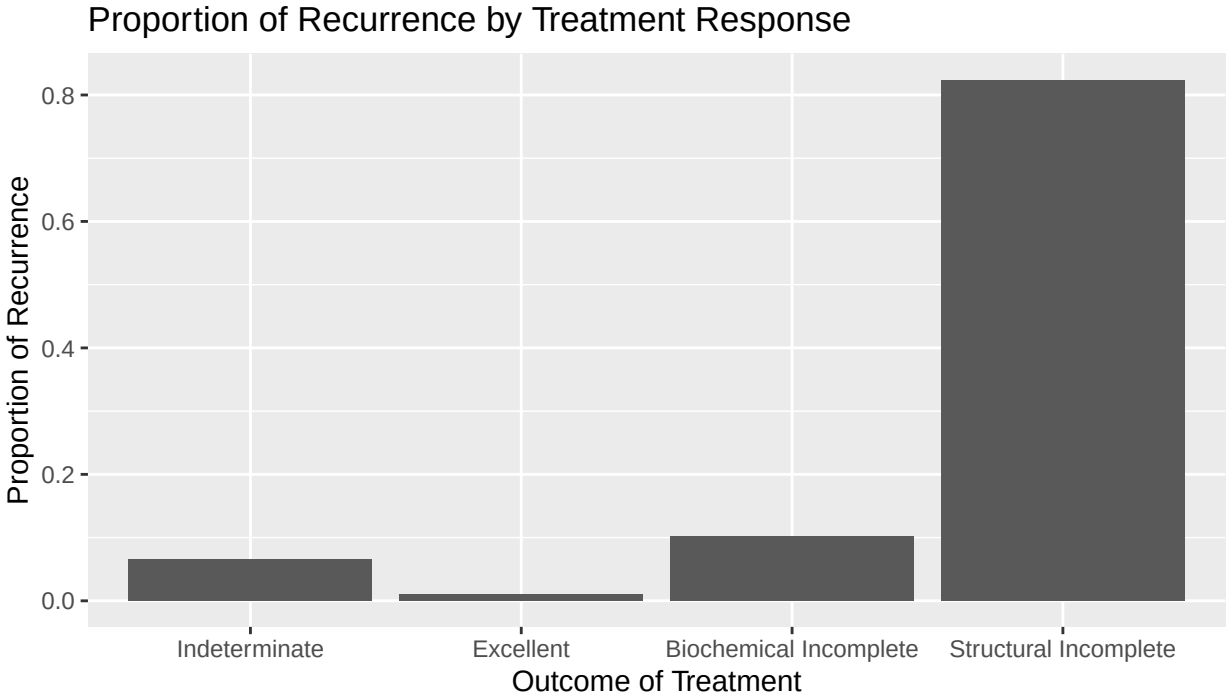
The fitted model is structured as follows, with Indeterminate as the reference category for the TreatmentOutcome variable:

$$\begin{aligned}\log\left(\frac{\pi}{1-\pi}\right) = 22.392 \\ &- 3.034\,(\text{TreatmentOutcomeExcellent}) \\ &+ 1.663\,(\text{TreatmentOutcomeBiochemicalIncomplete}) \\ &+ 5.671\,(\text{TreatmentOutcomeStructuralIncomplete}) \\ &+ 0.030\,(\text{Age}) \\ &+ 1.043\,(\text{GenderM}) \\ &- 10.634\,(\text{RadiotherapyYes}) \\ &- 26.058\,(\text{risk\_binaryNotHigh})\end{aligned}$$

Based on the table, we can see that TreatmentOutcome is a strong and statistically significant predictor of thyroid cancer recurrence. Excellent had a p-value of 0.005, Biochemical Incomplete had a p-value of 0.009, and Structural Incomplete had a p-value < 0.001. Since TreatmentOutcome is a categorical variable encoded via dummy variables, its coefficients can be interpreted without standardization.

Excellent outcome group had a coefficient of -3.034, which corresponds to an odds ratio of $e^{-3.034} = 0.048$. This means that their odds of recurrence are about 0.048 times those of the Indeterminate group, holding all other variables constant, indicating substantially lower odds of recurrence than those with an Indeterminate outcome. The Biochemical Incomplete outcome group has a coefficient of 1.663. The corresponding odds ratio is $e^{1.663} = 5.275$, indicating that their odds of recurrence are approximately 5.275 times greater than those in the Indeterminate group. The Structural Incomplete outcome group has a coefficient of 5.671. The odds ratio is $e^{5.671} = 290.32$, meaning that the Structural Incomplete group has approximately 290.32 times greater odds of recurrence than the Indeterminate group, reflecting dramatically increased odds of recurrence.

These results confirm that treatment outcome is a strong and statistically significant predictor of thyroid cancer recurrence. The direction and magnitude of the coefficients align with clinical expectations: the better the initial treatment outcome, the lower the risk of recurrence. The bar plot below visualizes the proportion of recurrence across treatment outcome categories.



While both Age and Gender were not statistically significant predictors at the conventiona 0.05 level (with p-value = 0.125 and p-value = 0.107 respectively), they are close to the significance level of 0.01 and may be worth investigating. Notably, holding all else constant, the odds of recurrence increases by approximately 3% per year of age, as $e^{0.03} = 1.03$. To investigate the individual contributions of Age and Gender, we conducted univariate logistic regressions using Age and Gender as the sole predictors to isolate individual effects.

Table 2: Univariate Logistic Regression: Recurred ~ Age

term	estimate	std.error	statistic	p.value
(Intercept)	-2.5227446	0.3550921	-7.10448	1.21e-12
Age	0.0373481	0.0076294	4.89525	9.82e-07

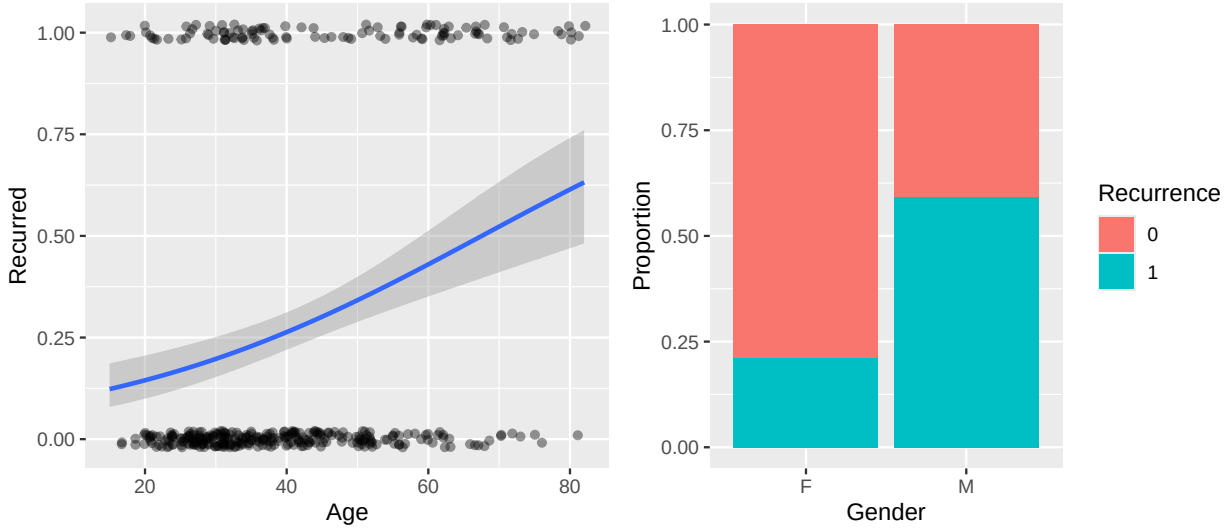
Table 3: Univariate Logistic Regression: Recurred ~ Gender

term	estimate	std.error	statistic	p.value
(Intercept)	-1.315677	0.1386237	-9.490993	< 2e-16
GenderM	1.686051	0.2784041	6.056128	1.39e-09

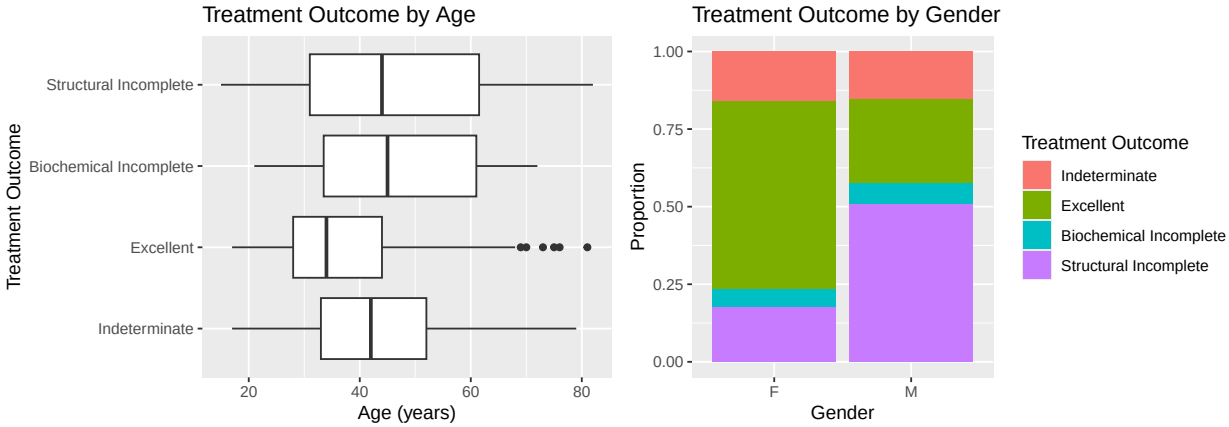
Interestingly, when modeled individually, both Age and Gender were statistically significant predictors of thyroid cancer recurrence. In the univariate model, Age had a p-value = 9.82e-07, indicating a significant positive association with recurrence risk. Similarly, Gender was significant in its respectivel model, with a p-value = 1.39e-09.

However, as we saw earlier, neither Age nor Gender were statistically significant at the 0.05 level in the multivariable model. This discrepency may be due to confounding or multicollinearity. That is, the effects of Age and Gender may be explained or diminished once other covariates, such as treatment outcome or clinical risk classification, are taken into account.

The exploratory plots below further support potential relationships between Age, Gender, and Recurrence. The logistic curve in the left panel suggests a positive association between Age and Recurrence, while the proportional bar chart in the right panel indicates that recurrence was more common among male patients.



We also examined whether Age and Gender are associated with the patient's treatment outcome through data visualizations.



In the boxplot comparing Age across Treatment Outcome categories, we observe that patients in the Excellent group tend to be younger, with a noticeably lower median age compared to the other groups. The remaining categories exhibit more similar age distributions with greater spread and higher medians. This suggests that younger age may be associated with more favorable treatment outcomes.

Gender also appears to influence treatment response. Female patients appear disproportionately more likely to achieve an Excellent treatment response, while male patients appear disproportionately more likely to be in the Structural Incomplete group. The Indeterminate and Biochemical Incomplete categories are more evenly distributed across genders.

0.4 FLOW WITH PLOT THOUGHTS?

The trends in Age were formally assessed using a one-way ANOVA. The conditions for this model are sufficiently satisfied to imply correlation, and are discussed in the appendix.

Table 4: One-Way ANOVA: Age by Treatment Outcome

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TreatmentOutcome	3	6357.015	2119.0050	9.897598	2.7e-06
Residuals	379	81141.194	214.0929	NA	NA

A One-Way ANOVA revealed statistically significant differences in Age across treatment outcome groups. We follow with a Tukey's post-hoc test to analyze the pairwise differences.

Table 5: Tukey HSD:Age by Treatment Outcome

Comparison	Difference	Lower CI	Upper CI	Adjusted p-value
Excellent-Indeterminate	-5.520	-11.017	-0.022	0.049
Biochemical Incomplete-Indeterminate	4.066	-5.172	13.305	0.668
Structural Incomplete-Indeterminate	3.438	-2.809	9.686	0.488
Biochemical Incomplete-Excellent	9.586	1.289	17.883	0.016
Structural Incomplete-Excellent	8.958	4.213	13.704	0.000
Structural Incomplete-Biochemical Incomplete	-0.628	-9.440	8.184	0.998

Post-hoc Tukey HSD tests revealed that the Excellent group were statistically significantly different from each of the other categories, reinforcing the visual trend observed in the boxplot. This confirms the inverse relationship between age and favorable treatment response.

Although this analysis is exploratory and descriptive in nature, it provides initial evidence that both Age and Gender may influence treatment response, which in turn may affect recurrence risk. Given these relationships, we considered the potential for multicollinearity, particularly between Age, Gender, and TreatmentOutcome.

Table 6: Generalized Variance Inflation Factors (GVIFs)

	GVIF	Df	GVIF^(1/(2*Df))
TreatmentOutcome	1.120	3	1.019
Age	1.105	1	1.051
Gender	1.032	1	1.016
Radiotherapy	2.068	1	1.438
risk_binary	2.068	1	1.438

Interestingly, adjusted GVIF values were well below the conventional thresholds of 5 ~ 10, indicating that multicollinearity is not a serious concern in this model. We also examined pairwise Pearson correlations among numeric predictors.

Table 7: Correlation Matrix of Numeric Predictors

	Age	Gender	Radiotherapy	Risk
Age	1.00	0.19	0.18	0.29
Gender	0.19	1.00	0.24	0.22

	Age	Gender	Radiotherapy	Risk
Radiotherapy	0.18	0.24	1.00	0.38
Risk	0.29	0.22	0.38	1.00

Radiotherapy and Risk exhibit a moderate correlation (0.38), which aligns with the slightly elevated GVIFs for both variables. This relationship is clinically plausible, as high-risk patients are more likely to receive radiotherapy. All other correlations are weak to moderate and do not indicate concern for multicollinearity.

0.5 Discussion

Our results support the hypothesis that treatment outcome is a strong and statistically significant predictor of recurrence. Specifically, patients with a structural incomplete response had dramatically elevated odds of recurrence, while those with an excellent response had substantially lower odds.

Although Age and Gender were not statistically significant in the multi logistic regression model, univariate and exploratory analyses suggested that they are individually associated with recurrence and may influence treatment outcome.

However, these findings must be interpreted within the context of the study's limitations. For one, the data was not randomly sampled from the broader thyroid cancer population; it came from a single medical center. Although each patient was independently observed, the sample lacks random representation of the general population. This limits the extent to which we can generalize our conclusions to all patients with thyroid cancer.

Additionally, because this was an observational study, causal inference cannot be made. We can say that treatment outcome is strongly associated with recurrence, but we cannot conclude that one causes the other. It is possible that unmeasured confounding variables (e.g. genetic markers, comorbidities) may influence both treatment response and recurrence risk.

Risk and Radiotherapy also had large standard errors due to uneven distributions. Though not statistically significant predictors in this model, their relationship remains clinically important and warrants cleaner data or larger sample sizes to clarify their roles.

Potential selection bias may also be present, as the dataset may overrepresent patients with more complete follow up cases, given the 15-year tracking requirement.

Future research should consider including additional clinical variables such as tumor size, molecular markers, or comorbidities that may help refine prediction. Further studies should also consider experimental designs to enable causal inference, and use more diverse samples for generalizability.

0.6 References

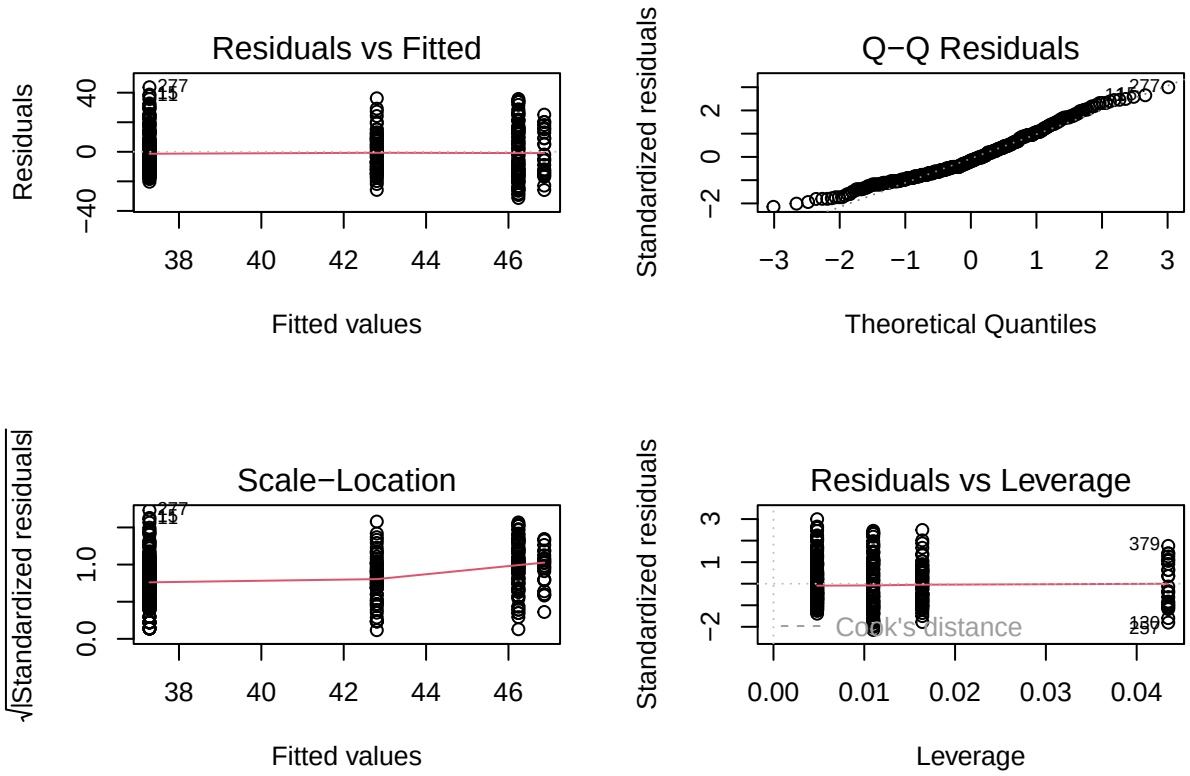
Borzooei, S., Briganti, G., Golparian, M. et al. Machine learning for risk stratification of thyroid cancer patients: a 15-year cohort study. Eur Arch Otorhinolaryngol 281, 2095–2104 (2024). https://doi.org/10.1007/s00405-023-08299-w.

"Common Cancer Sites - Cancer Stat Facts." SEER (2025). seer.cancer.gov/statfacts/html/common.html#comparison.

"Thyroid Cancer." Mayo Clinic, Mayo Foundation for Medical Education and Research, 5 Jan. 2024, www.mayoclinic.org/diseases-conditions/thyroid-cancer/symptoms-causes/syc-20354161.

"Thyroid Cancer Recurrence Dataset." Kaggle (2025). https://www.kaggle.com/datasets/aneevinay/thyroid-cancer-recurrence-dataset

0.7 Appendix



#! DO WE NEED THIS The diagnostic plots show that the ANOVA model conditions have been relatively well satisfied. Randomness and Independence have been decently met as discussed in the Methods section. In the Residuals vs Fitted plot, the spread of the residuals appears to be relatively consistent across the range of fitted values. There isn't a clear indication of increasing or decreasing variance. In the Normal Q-Q Plot, the residuals appear to be normally distributed. There are some minor deviations at the tails, but they don't appear to be drastically severe. In the Scale Location plot, the points seem to be scattered relatively randomly around a horizontal line. There isn't a strong trend suggesting unequal variance. The Residuals vs Leverage plot also doesn't indicate the presence of highly influential outliers that would severely violate the ANOVA assumptions.