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. These cells can shunt blood flow to maintain their high metabolic cost and increasing size (potentially blocking organ function), which becomes deadly, especially if it metastasizes (spreading elsewhere in the body) (Mayo Clinic 2024). In the United States, cancer is the leading cause of death for people under 65 years old (Cancer.gov 2025). We plan to study thyroid cancer specifically, which is projected to have 44,000 new cases this year, and is unique in that (although statistically insignificant) its death rate has trended upwards in recent years (whereas many have had a statistically significant decrease) (Cancer.gov 2025). Thyroid cancer occurs in the thyroid, an endocrine system gland located at the base of the neck that regulates heart rate, blood pressure, body temperature and weight hormonally (Borzooei et al. 2023).

Luckily, thyroid cancer has treatment options, including surgical removal and Radioactive Iodine therapy (RAI) that targets cancerous thyroid cells by exposing them to radioactive iodine (which is primarily taken up by the thyroid) (Mayo Clinic 2024). However, treatment response can vary, and monitoring how well a patient responds is critical. We believe that a poor response may increase the likelihood of recurrence. As such, understanding factors such as treatment outcome on recurrence is highly valuable for deciding a patient's next steps.

This study investigates the following research question: Holding other relevant factors constant (age, gender, prior radiotherapy, and clinical risk classification), is initial treatment response a statistically significant predictor of thyroid cancer recurrence? We hypothesize that the response to the treatment 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 relationships among these explanatory variables for further statistical analysis.

0.2 Methods

The dataset for this study 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 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

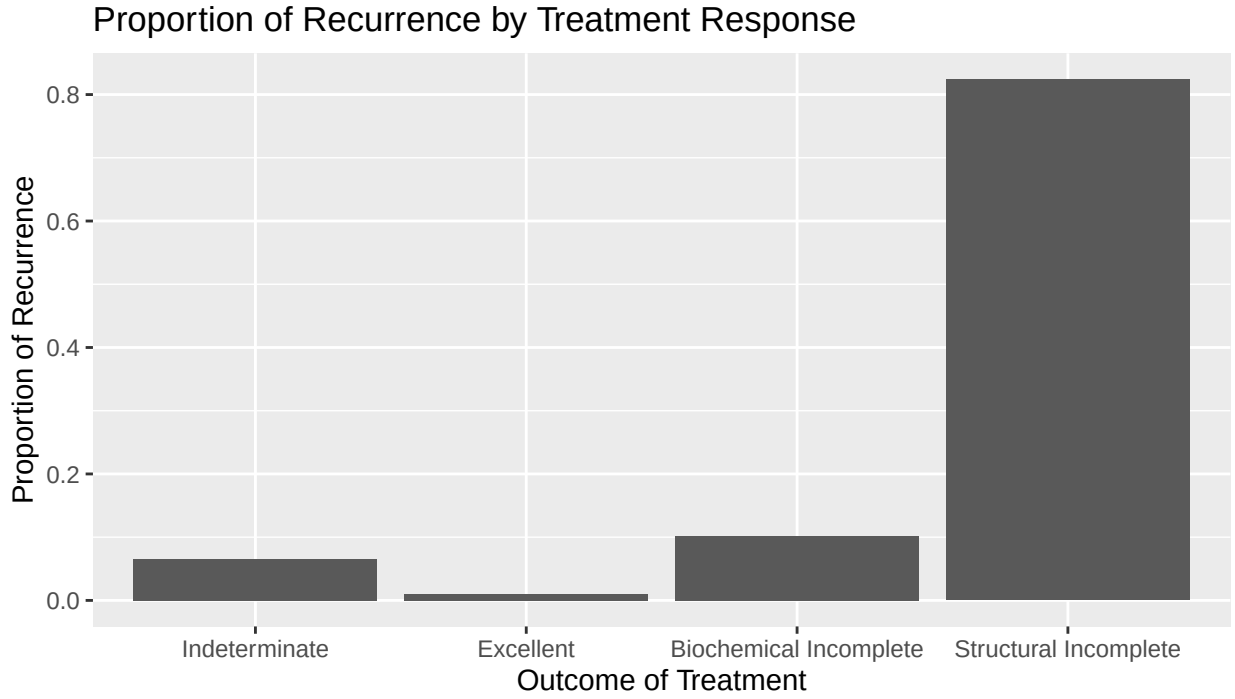
The fitted model is structured as follows, with Indeterminate as the reference category for the TreatmentOutcome variable:

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

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. #! CAN WE INTERPRET COEFFICIENTS EVEN IF WE DIDNT STANDARDIZE CUZ CATEGORICAL AND USED DUMMY?? 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.

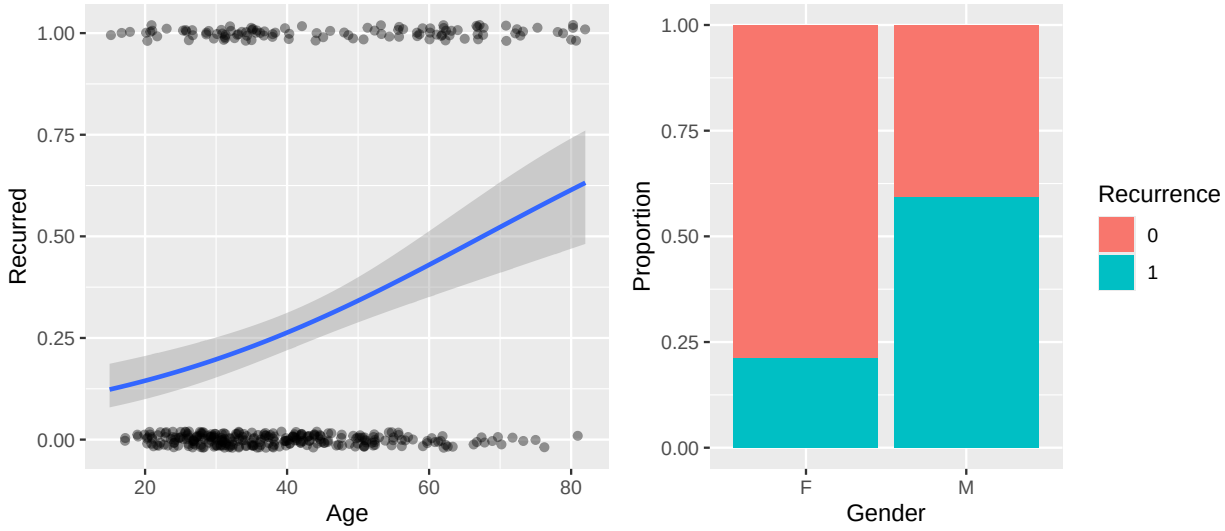
The odds ratio is $e^{5.671} = 290.32$, meaning that the Structural Incomplete group has approximately 290.32 times greater odds of recurrence than the Indeterminate group, reflecting dramatically increased odds of recurrence.

These results confirm that treatment outcome is a strong and statistically significant predictor of thyroid cancer recurrence. The magnitude of these odds ratios suggest that the clinical response to initial treatment plays a crucial role in long-term recurrence risk. The bar plot below visualizes the proportion of recurrence across treatment outcome categories.

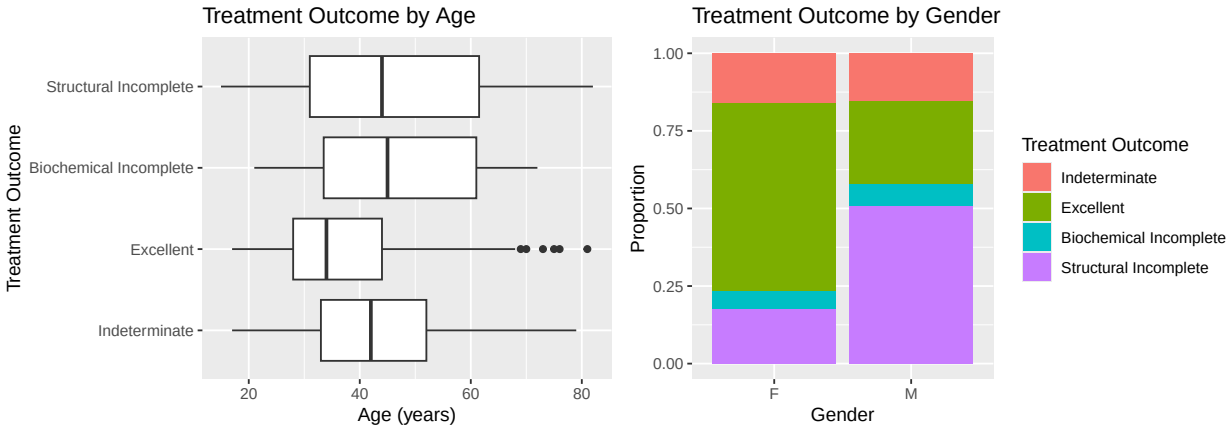


While both Age and Gender were not statistically significant predictors at the conventiona 0.05 level (with p-value = 0.125 and p-value = 0.107 respectively), they are close to the significance level of 0.01 and may be worth investigating. Notably, holding all else constant, the odds of recurrence increases by approximately 3% per year of age, as $e^{0.03} = 1.03$. To explore these relationships further, we conducted univariate logistic regressions (see Appendix), where both Age and Gender were found to be individually significant predictors of recurrence. This discrepancy may be due to multicollinearity or overlapping variance with TreatmentOutcome.

The exploratory plots of recurrence by Age and Gender below further support their potential relevance.



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



In the boxplot comparing Age by Treatment Outcome categories, we observe that the Excellent outcome group has a visibly low median age compared to other groups, while the remaining groups show similar distributions. Gender also appears to influence treatment response. The stacked bar chart shows that females are disproportionately more likely to have an Excellent outcome. Conversely, Males are disproportionally much more likely to get a Structural Incomplete treatment outcome than Females. Indeterminate and Biochemical Incomplete outcomes appear relatively evenly distributed across genders.

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 2: 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 3: 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 4: Generalized Variance Inflation Factors (GVIFs)

	GVIF	Df	GVIFˆ(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

All 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 5: 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
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 clincially 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.

0.4 Discussion

This study investigated whether initial treatment response is a statistically significant predictor of thyroid cancer recurrence, controlling for age, gender, radiotherapy history, and risk classification. Our analysis revealed hat 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. These findings support our hypothesis and highlight the

importance of monitoring and interpreting treatment response when planning long-term care for thyroid cancer patients.

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

While our findings offer meaningful insights, they are primarily generalizable to the sample population from which the data were drawn. The data was not randomly sampled from the broader thyroid cancer population, which limits the extent to which we can generalize our conclusions to all patients with thyroid cancer. The assumptions of independence and randomness were only partially satisfied. Although each patient was independently observed, the sample lacks random representation of the general population. 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. Further studies using randomized controlled trial data or longitudinal designs would be better suited for identifying causal relationships.

There are several limiations to this study. For one, Risk and Radiotherapy had large standard errors. Though not statisticaly 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.

0.5 Appendix

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.

Source Used for Data:

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

Data:

```
#Both Age and Gender individually are statistically significant
logistic_Age <- glm(Recurred ~ Age, data, family = "binomial")
summary(logistic_Age)
```

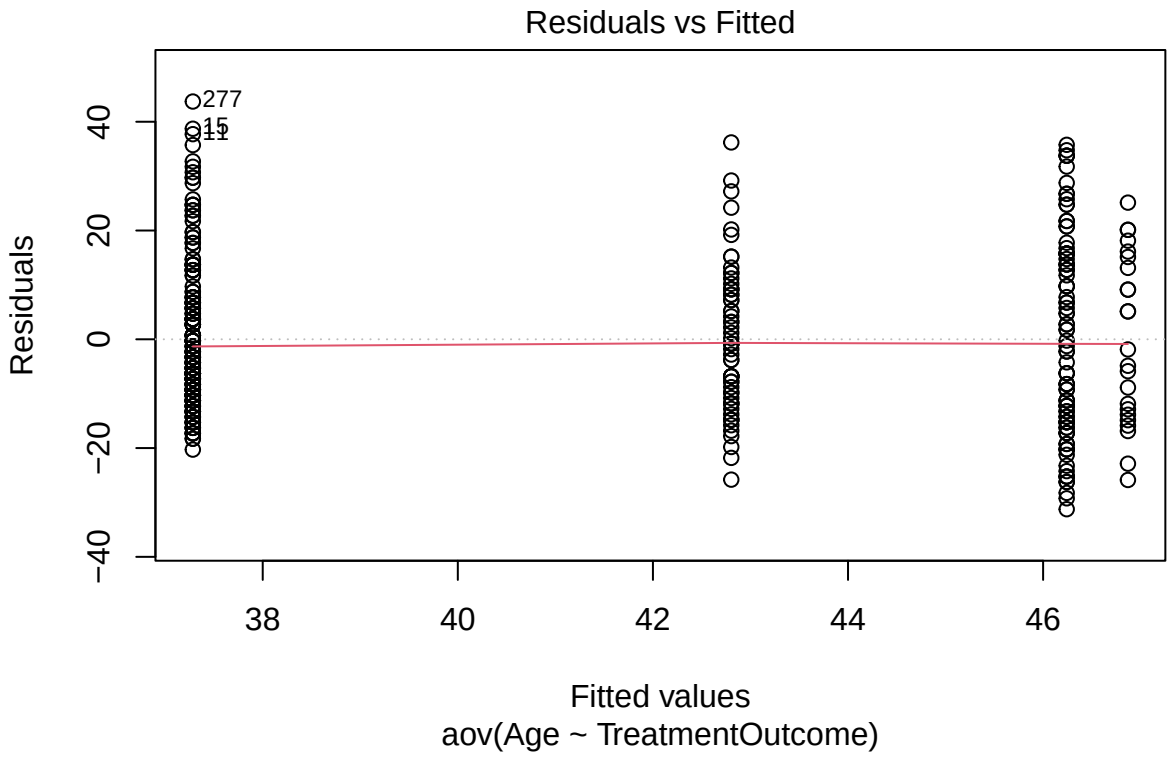
```
##
## Call:
## glm(formula = Recurred ~ Age, family = "binomial", data = data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.522745   0.355092  -7.104 1.21e-12 ***
## Age          0.037348   0.007629   4.895 9.82e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 455.63  on 382  degrees of freedom
## Residual deviance: 430.58  on 381  degrees of freedom
## AIC: 434.58
##
## Number of Fisher Scoring iterations: 4
```

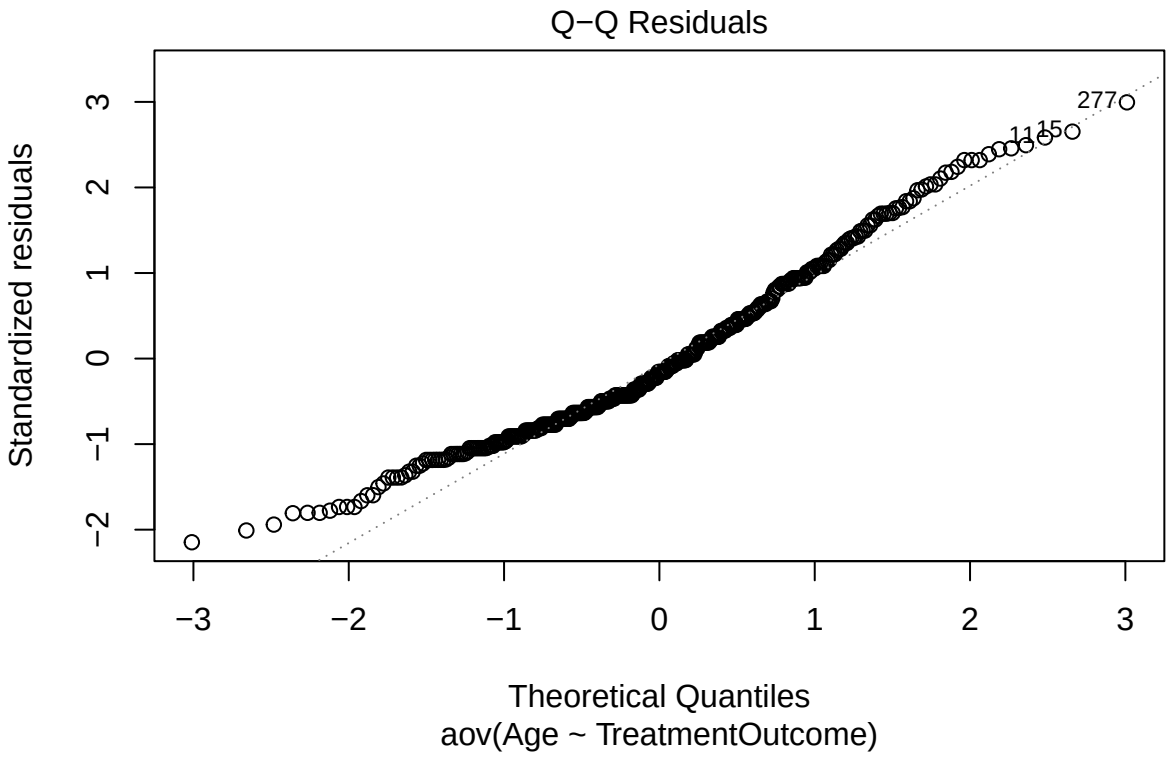
```
logistic_Gender <- glm(Recurred ~ Gender, data, family = "binomial")
summary(logistic_Gender)
```

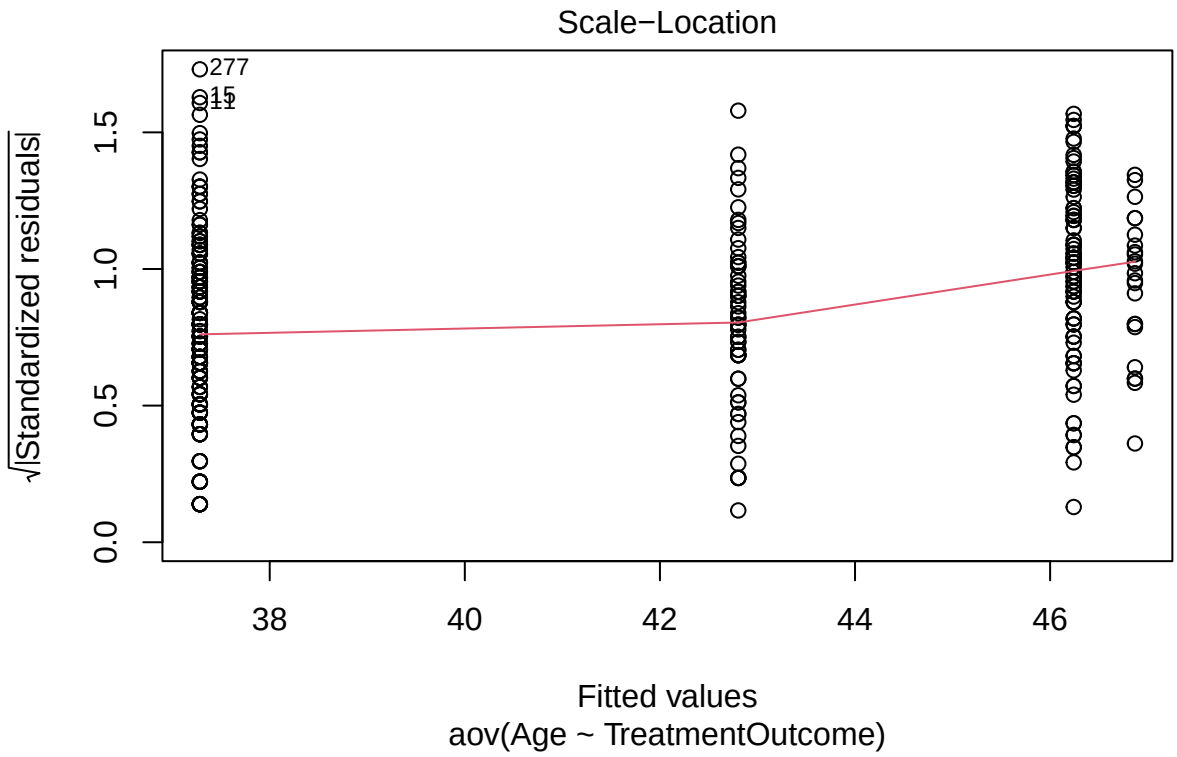
```
##
## Call:
## glm(formula = Recurred ~ Gender, family = "binomial", data = data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.3157     0.1386  -9.491  < 2e-16 ***
## GenderM       1.6861     0.2784   6.056 1.39e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

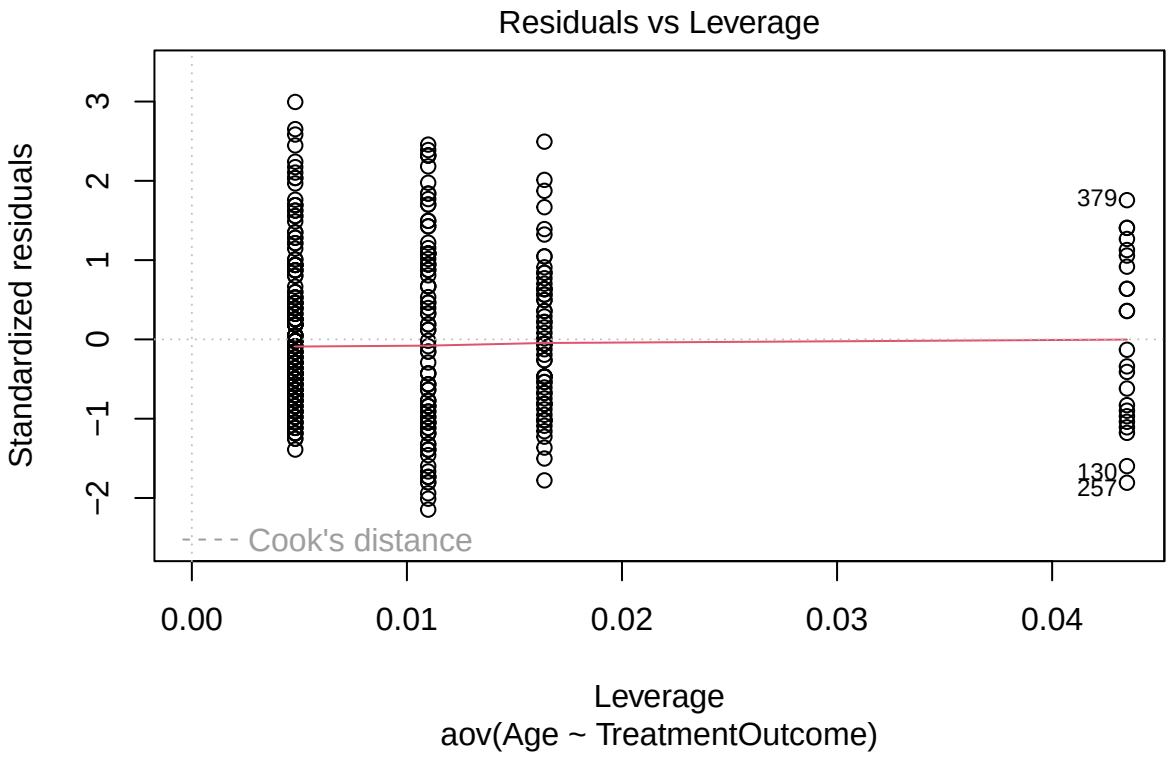
```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 455.63  on 382  degrees of freedom
## Residual deviance: 418.01  on 381  degrees of freedom
## AIC: 422.01
##
## Number of Fisher Scoring iterations: 4
```

```r
#Discuss the ANOVA model conditions #!
plot(model)
```




Q−Q Residuals
Standardized residuals
3
2
1
0
−1
−2
277
115
111
−3
−2
−1
0
1
2
3
Theoretical Quantiles
aov(Age ~ TreatmentOutcome)


Scale−Location
277
15
11
√|Standardized residuals|
0.0
0.5
1.0
1.5
38
40
42
44
46
Fitted values
aov(Age ~ TreatmentOutcome)



We see that this satisfies the ANOVA model conditions well enough to imply correlation:

Residuals are random and independent - discussed in the Methods section (the logarithmic analysis for this condition is the same here)

Effects are constant & additive - Residuals vs fitted graph

Residuals have the same variability in each group - based on the boxplot shown in the analysis, this seems relatively satisfied, only the excellent condition has a lower variability

Residuals are normally distributed - the residuals deviate from the normality line quite drastically at lower theoretical quantiles