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). In the United States, cancer is the leading cause of death for people under 65 years old. 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).1 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.

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). 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 response on recurrence is highly valuable for deciding a patient's next steps, which is what we plan to study.

Our research question is as follows: Holding other relevant factors (age, history of radiotherapy, adenopathy, and clinical risk classification) constant, how does initial treatment response predict thyroid cancer recurrence?

0.2 Methods

Our data was obtained from kaggle, which had been collected from an article submission from Hamadan University, published in the European Archives of Oto-Rhino-Laryngology. The data collected was from 383 individual thyroid cancer patients, whose data was collected over 15 years, and each patient was followed for at least 10 years. The information collected was about variables that relate to the patient's family history, treatment history, and outcomes

To answer our research question, we will examine the following variables:

- Recurrence (Binary Categorical Dependent Variable, whether or not cancer recurred)
- Age (Quantitative Explanatory Variable, age of the patient in years)
- Radiotherapy (Binary Categorical Explanatory Variable, history of prior radiotherapy)
- Adenopathy (Binary Categorical Explanatory Variable, presence of lymph node involvement)
- Risk (3 Level Categorical Explanatory Variable, cancer risk classification categorized as low, medium, high)
- Response (4 Level Categorical Explanatory Variable, treatment response classified as excellent, indeterminate, structural incomplete, biochemical incomplete).

```
data$Recurred <- fct_recode(data$Recurred,
                     "1" = "Yes",
                     "0" = "No")
```

```
logistic_model <- glm(Recurred ~ Age + Hx.Radiothreapy + Adenopathy + Response,
                     data, family = binomial)
summary(logistic_model)
```

```
##
## Call:
## glm(formula = Recurred ~ Age + Hx.Radiothreapy + Adenopathy +
```

```
##      Response, family = binomial, data = data)
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -1.25748    1.46080  -0.861 0.389339
## Age                          0.03843    0.02009   1.913 0.055741 .
## Hx.RadiothreapyYes           2.21588    5.10264   0.434 0.664099
## AdenopathyExtensive         13.38688 1399.62488   0.010 0.992369
## AdenopathyLeft               1.06417    1.41845   0.750 0.453112
## AdenopathyNo                -0.96115    0.91789  -1.047 0.295037
## AdenopathyPosterior         13.59693 2778.46207   0.005 0.996095
## AdenopathyRight              0.36319    1.08735   0.334 0.738370
## ResponseExcellent           -4.85976    1.11105  -4.374 1.22e-05 ***
## ResponseIndeterminate       -2.02442    0.64314  -3.148 0.001646 **
## ResponseStructural Incomplete 3.46555   0.91232   3.799 0.000146 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 455.630  on 382  degrees of freedom
## Residual deviance:  96.411  on 372  degrees of freedom
## AIC: 118.41
##
## Number of Fisher Scoring iterations: 16
```

Below is our multiple logistic regression model:

0.3 Analysis and Results

0.4 Discussion

0.5 References