P53 Project Report: Determining an Appropriate Threshold for Immunohistochemistry (IHC) Staining Test

Alyssa Rogers-Armstrong

February 25th 2024

Executive Summary

The objective of this study was to evaluate the appropriateness of the 10% threshold for predicting TP53 mutation status in brain cancer patients based on IHC staining percentages. Statistical analysis indicates that the 10% threshold is not an accurate predictor of mutation status and instead we should adopt a 40% threshold. The 40% threshold has better performance in terms of sensitivity, specificity, and AUC. By adopting this new threshold, the accuracy and reliability of the TP53 mutation identification is enhanced, which leads to fewer instances of false positives (incorrectly identifying patients with a mutation who do not have it) and false negatives (incorrectly identifying patients without a mutation who do have it).

Introduction

The accurate identification of genetic mutations, particularly in the TP53 gene, is crucial in oncology for understanding tumor biology and treatment decisions. In the context of brain cancer, TP53 mutations are prevalent. Cancer researchers have noted that the TP53 gene is mutated in more than 50% of all cancers. The detection of TP53 mutations in brain cancer patients can provide valuable insights into the specific type of cancer a patient has, therefore leading to a more targeted approach to the treatment the patient receives.

While the gold standard in the industry is to use the Sanger Sequencing method, this method is a highly accurate but expensive way to classify mutation status. We want to investigate the use of the immunohistochemistry (IHC) staining test to identify mutation status instead. This method is less expensive and more accessible than the Sanger Sequencing method. Currently, the industry recommends patients to undergo further testing and treatment is the IHC staining test reports 10% of the nuclei stained or higher, however, this threshold appears to be problematic.

In this study, we aim to evaluate the 10% threshold and propose a new threshold based on statistical analyses which must:

- Accurately predict TP53 mutation status, without too many false positives or too many false negatives
- Be protective against false negatives since this would be disastrous to patients that have a GMB (Glioblastoma Multiforme)

Methodology

Researchers collected a total of 41 tissue samples from patients with brain cancer being treated in the Roanoke, Virginia area. These samples were stained with the compound IHC and the percent of nuclei staining positive was recorded for each patient. Additionally, the presence or absence of a TP53 gene mutation was determined using the Sanger Sequencing. The results of the Sanger Sequencing and the IHC staining test are recorded in the file p53.csv. All of the data points were used in the analysis.

Statistical analyses were conduced using R statistical software. Logistic regression models were fitted to assess the relationship between IHC staining percentages and TP53 mutation status. Sensitivity, specificity, and AUC were calculated to evaluate different threshold performances.

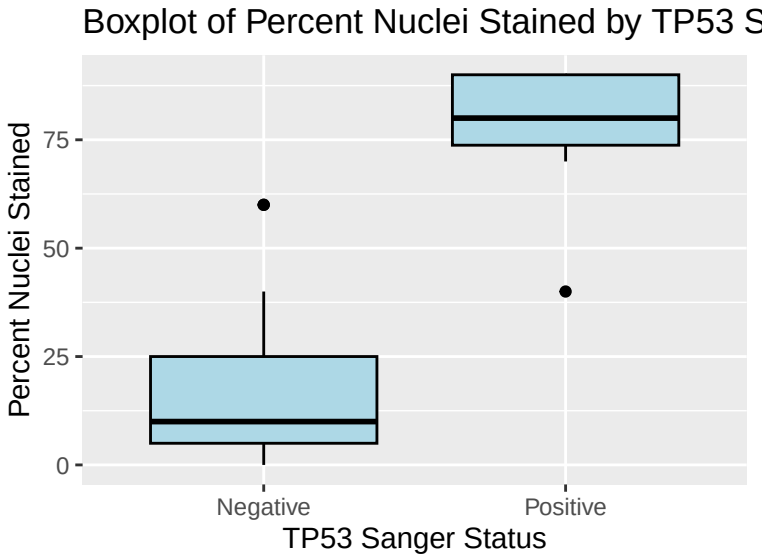
Analysis

Visualizations were created to facilitate the interpretation of the analysis results. Logistic regression modeling was completed to analyze the association between the percentage of nuclei staining positive in IHC and the probability of TP53 mutation. Logistic regression is a statistical technique used for modeling the relationship between a binary outcome variable and one or more predictor variables. The logistic regression model was then used to obtain odds ratios. Odds ratios provide insights into the strength and direction of the relationship between predictor variables and the outcome. Additionally, sensitivity, specificity, and AUC (area under the curve) were calculated to compare the 10% and 40% thresholds.

Results

For this study, statistical analysis was conducted to investigate the relationship between the TP53 gene mutation status (found by Sanger Sequencing) and the percentage of nuclei stained by IHC staining. The goal was to investigate the 10% threshold of stained nuclei and to see if a better threshold would fit the data. This would improve the accuracy in which patients are diagnosed with a mutation and will assist with the treatment options in patients with brain cancer.

Figure 1 below provides a visualization of our data by positive and negative status of a mutation.



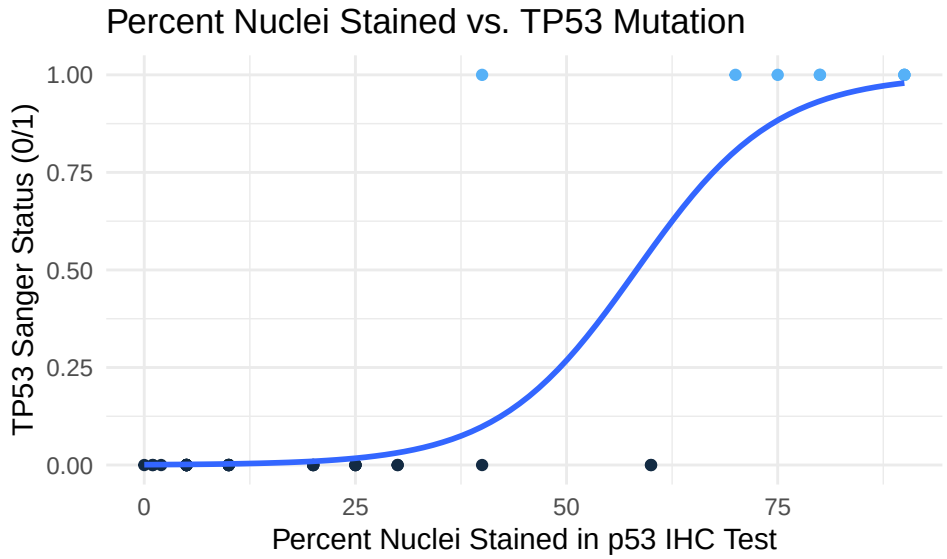
Several insights can be gained from Figure 1:

- There are two outliers in the data, one from the negative classifications and one from the positive ones.
- The TP53 Sanger Status with no mutations (negative) appears to have a lower mean of nuclei stained than the TP53 Sanger Status with mutations (positive).
- The variability or the spread of the data for each group appears to be approximately equal.

For the analysis, logistic regression analysis was performed first. Logistic regression analysis is a statistical technique used to model binary outcomes (yes and no, positive and negative, etc.). Additionally, logistic

regression models will estimate the probability of an event occurring based on one or more variables. A logistic regression model was fit to the data using the glm() function in R. From this function, a p-value of 0.006 was obtained which is statistically significant due to it being smaller than 0.05. This means that we can say that a one unit increase in Percent of Nuclei Stained is associated with an average increase of 0.12120 in the log odds of TP53 Sanger Status.

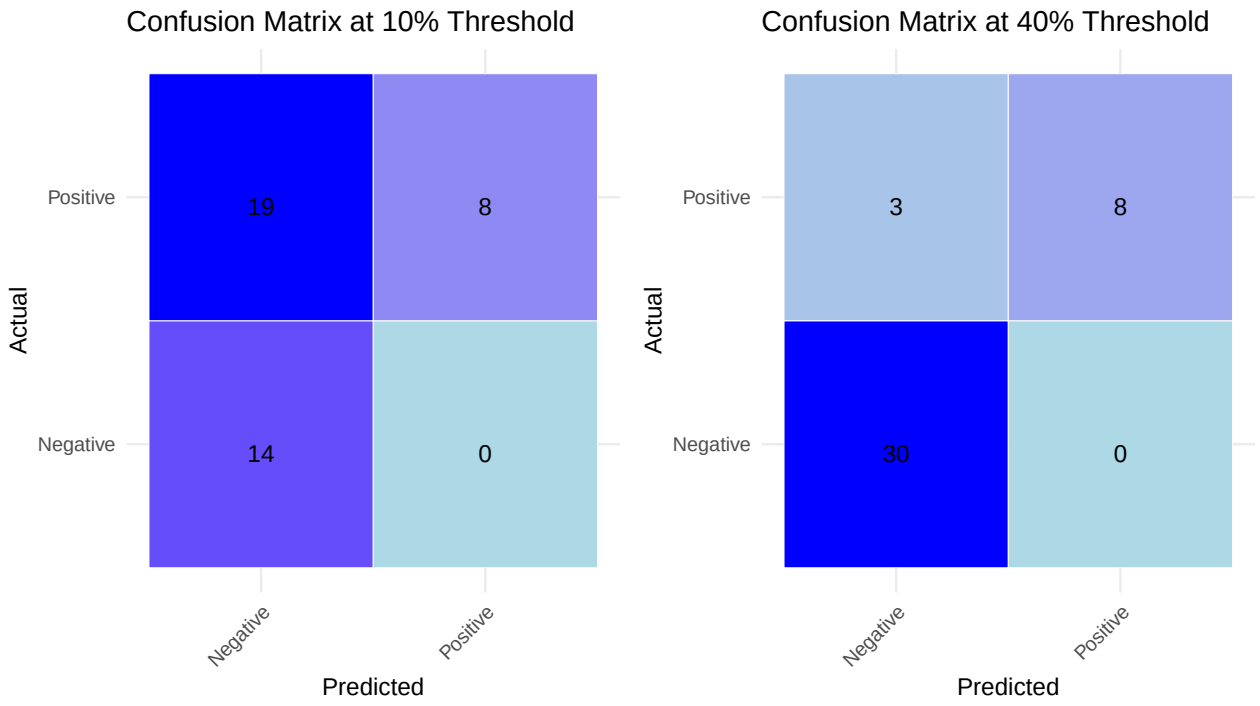
Figure 2 below provides a visualization of our individual data with the logistic regression curve displayed over the data.



The odds ratio was calculated from the logistic regression model. Odds ratios are used to quantify the strength and direction of the association between the predictor variable (percentage of stained nuclei) and the odds of having a TP53 mutation. The odds ratio was calculated to be 1.12885. This means that for every one-unit increase in the percentage of nuclei stained with p53 in IHC, the odds of having a TP53 mutation increase by approximately 12.9%. Odds ratios above 1 indicate that there is a positive association between the extent of p53 staining in IHC and the presence of TP53 mutations in brain cancer patients. This helps to provide insight in the use of IHC staining as a diagnostic tool for identifying TP53 mutations in brain cancer patients. However, the 95% confidence interval for the odds ratio includes 1. This suggests that the association between TP53 mutations and the IHC staining may not be statistically significant.

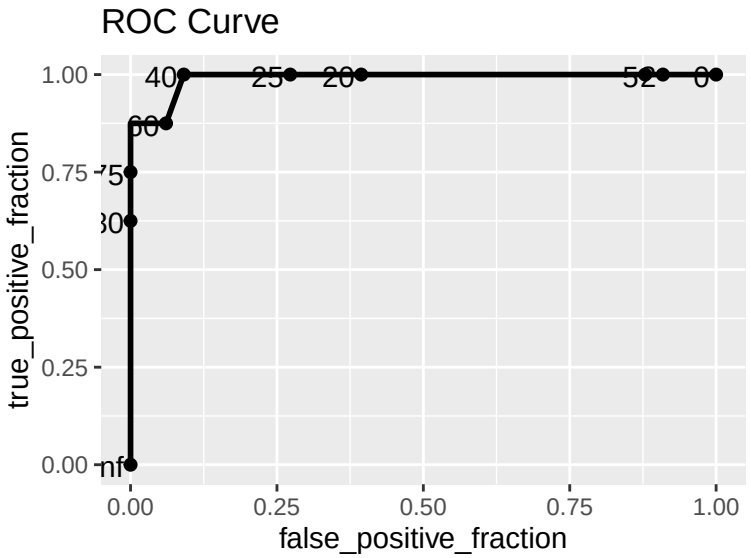
After determining from the model that IHC staining is a good diagnostic tool for identifying TP53 mutations in brain cancer patients, the thresholds were evaluated to see which would be the best fit for the data. The data was evaluated at the 10% threshold first and then a 40% threshold. A 40% threshold was chosen using the cutpointr function. The cutpointr function in R determines the optimal threshold for continuous data based on the association with a binary outcome variable.

Figure 3 below provides a visualization of the confusion matrix results.



The confusion matrix at the 10% threshold has a sensitivity of 1 and a specificity of 0.4242. This means that at the 10% threshold, the test correctly identifies all patients who have the mutation but it only identifies 42% of patients who do not have the mutation. This means that there are 19 patients who are being identified as having the mutation and are undergoing further testing, but they do not actually have the mutation. The confusion matrix at the 40% threshold has a sensitivity of 1 and a specificity of 0.9091. This means that at the 40% threshold, the test correctly identifies all patients who have the mutation but only identifies 91% who do not have the mutation. This means there are 3 patients who are being identified as having the mutation and are undergoing further testing, but they do not actually have the mutation. Using the 40% threshold leads to less false positive results, and therefore, results in less money wasted on unnecessary further testing.

In addition to using the confusion matrix and cutpointr function to determine the best threshold, the ROC curve was looked at. Figure 4 below shows the ROC curve of the data.



This curve represents the probability that the model will rank a randomly chosen positive instance higher than a randomly chosen negative instance. In this graph we can see the trade-off between sensitivity and specificity. In this study, it is best to have a high sensitivity so that we are not encountering false negatives. This could be disastrous to the patient who receives them. It is also good to have as high of a specificity without compromising our sensitivity. The ROC curve shows that this would be at the 40% threshold. From this graph the AUC (area under the curve) can be calculated. The AUC values range from 0 to 1. The higher values indicate better discriminatory power. For this ROC curve, the AUC is calculated to be 0.9905.

Interpretation and Recommendations

The threshold of 10% staining percentage for determining TP53 mutation status is arbitrary. Analysis of the data suggests that the 10% threshold may lead to misclassification of mutation status resulting in false positives or false negatives. It is essential to acknowledge the limitations of the data. This data set does not include demographic information, such as age, gender, etc. Without having access to this information, it is difficult to make assumptions beyond the scope of the data analysis.

The recommendations are as follows:

- Threshold Reevaluation: It is recommended to utilize the 40% threshold instead of the 10% threshold. The 40% threshold showed better specificity, accuracy, and AUC when comparing it to the 10% threshold. This will ensure that less materials are wasted on patients that do not have the mutation and that patients are being correctly identified with the mutation. Overall, this leads to patients receiving more specialized care depending on their tumor type.
- Sample Size: In future studies, there should be a change in the sample size. Larger sample sizes enhance the reliability of the study and will allow for more precise estimation of parameters. The sample size is 41 in this study, this makes it difficult to ensure adequate statistical power for detecting associations between the TP53 mutation status and IHC staining percentages.
- Data Validation: Validation studies are important to understand the robustness and generalizability of the 40% threshold for using TP53 mutation status based on IHC staining percentages. In the future, independent validation studies should be conducted.
- Data Collection: In this study, the percentage of nuclei stained was obtained by an individual counting the nuclei stained in the brain cancer images once. In order for this study to be replicated in the future, a consistent process of counting the stained nuclei would need to be put in place and multiple people should count the stained nuclei images to ensure that the percentage number in the data is accurate and reliable.