

Ari Li
Undergraduate
Yale University, New Haven, CT

Mentor: Dr. Andrew G. Chapple, PhD
Stanley S Scott Cancer Center, Department of Interdisciplinary Oncology

“Are relationships between patient demographics and cancer stage different based on Gastrointestinal cancer site locations?”

Research has shown that the early detection of cancer through screening is linked with higher rates of survivability and a lower cost of care, among other benefits. However, there are many factors associated with the early detection of cancer including: age, diagnosis year, sex, insurance type, poverty, race, smoking, and location (urban vs rural).

Our goal is to determine whether the relationship between the factors listed and stage at diagnosis differed by cancer site. Data collected through the Louisiana Tumor Registry was used for this research project. The data collected between 2000 and 2020 contained over 15000 patients with one primary cancer site in their lifetime. We analyzed data from the gastrointestinal cancers with $n > 1000$ patients, including Ascending Colon, Cecum, Esophagus, Liver, Pancreas, Rectum, Sigmoid, Colon, and Stomach cancers.

We used the linear regression function in R Statistical Software (`lm`) to model the predicted stage at diagnosis given the listed factors and the cancer site. We also fit a linear regression model with an interaction between cancer site and each factor of interest (i.e. age). This baseline model was then compared using a nested model ANOVA test. This comparison was done separately for each factor of interest.

Of the interaction models generated for each factor, only the age at diagnosis, diagnosis year, and the insurance type returned statistically significant results ($p < 0.05$) from the ANOVA. Of these significant variables, each site-specific interaction coefficient from the interaction model was then compared against each other to determine the differences in association by cancer site. To do this, the estimated variance-covariance matrix of the interaction model was generated using the `vcov` function in `r`, and p-values were generated for testing equality of each pair of interaction coefficients using a chi-squared distribution. The significant coefficient comparisons were shown graphically via a heat map with colors displaying the difference in association. We found that for the age-predicted stage interaction model, pancreas and liver cancer did not decrease as the age of the patients increased in age, signaling a similar level of screening between older and younger patients. This differed from the negative trend (i.e. stage decreased with age) for all other cancer sites. We also found that for the diagnosis year-predicted stage interaction model, pancreas cancer's predicted stage did not decrease over the years, while the other GI cancers decreased (especially liver) drastically, indicative of a lack of improvement in pancreas cancer screening over the years. Lastly, we also found that for the insurance-predicted stage interaction model, cancer was detected earlier in those with private health insurance than those with public insurance. For pancreatic cancer this difference was significantly larger than all other cancer sites.