Can we Predict Rectal Cancer Outcomes using Clinical Data? A Comparative Analysis of Different Techniques.

Karina Krishnan, Grade 11, Beachwood High School | *Beachwood, Ohio, 44122*

Background and Research Question

Rectal cancer forms in the tissues of the rectum. Cancer inside the rectum and cancer inside the colon are often referred to together as colorectal cancer (CRC). CRC is the third most common cancer and the second leading cause of cancer-related deaths in the world with an estimated number of 1.8 million new cases and about 881,000 deaths worldwide in 2018. The objective of this study is to determine which pre-surgery and Magnetic Resonance Imaging (MRI) variable(s) are significantly associated with rectal cancer outcomes –measured by the pathologic T-Stage, pathologic N-Stage, pathologic M-Stage, under the TNM (Tumor, Node, Metastasis) system that describes the extent of the primary tumor based on its size and invasion into surrounding tissues, for patients diagnosed with locally advanced rectal cancer, or recurrence. This is important for prognostic assessment, surgical planning, and the evaluation of treatment response, in the context of multidisciplinary rectal cancer care.

Methodology

- Three different regression techniques are utilized to determine whether any variable of the variables is consistently and significantly associated with outcomes. Since the dependent variables are not continuous variables, Tobit regression is used for path_t_stage, path_n_stage, and path_m_stage which can take ordered values, and the Logit regression is used for Recurrence that is 0 or 1 indicator variable. LASSO and Ridge regression methods, as well as a combination of the two, called ElasticNet regression, are also used.
- Using Stata and Python programming, the regression results are examined with (Panel A of each table) only the imaging variables and (Panel B of each table) with both imaging variables and pre-surgery variables (or control variables).

Data Analysis and Results

Table 1: The results show that number_of_positive_lymph_n and lymphovascular_invasion imagining variables are significantly associated with path M Stage outcome in both Panels A and B.
Table 2: Results show that number_of_positive_lymph_n imagining variable is significantly associated with path T Stage and path N Stage outcomes in both Panels A and B. Among the pre-surgery or control variables, init_clinical_staging_m appears to be a significant predictor of path T Stage, path N Stage and path M Stage outcomes, and race appears to be a significant predictor of path N Stage and path M Stage outcomes in Panel B.
Table 3: Results show that number_of_positive_lymph_n imagining variable is significantly associated with path N Stage both Panels A and B.

***Table 1*: Tobit and Logit regression *Table 2*: Adaptive LASSO regression, SCAD, and MCP**

Panel A:

Significant Coefficients for Path T Stage	*Tobit*
mucin_present	**2.879**

Significant Coefficients for Path M Stage	*Tobit*
number_of_positive_lymph_n	**0.274**
lymphovascular_invasion	**3.314**

Panel B:

Significant Coefficients for Path M Stage	*Tobit*
number_of_positive_lymph_n	**0.572**
lymphovascular_invasion	**0.393**
sex	**-4.353**

Panel A:

Non-Zero Coefficients for Path T Stage	*Adaptive Lasso*	*SCAD*	*MCP*
number_of_positive_lymph_n	**0.073**	**0.036**	
distance_to_proximal_margin		-0.031	
number_of_lymph_nodes_exam		0.044	
Non-Zero Coefficients for Path N Stage	*Adaptive Lasso*	*SCAD*	*MCP*
number_of_positive_lymph_n	**0.214**	**0.156**	**0.090**

Non-Zero Coefficients for Path M Stage	*Adaptive Lasso*	*SCAD*	*MCP*
number_of_positive_lymph_n	0.044		
distance_to_proximal_margin		-0.034	
distance_to_distal_margin		-0.022	
number_of_lymph_nodes_exam		-0.045	

A: no significant lymph nodes;
B: presence of a lymph node

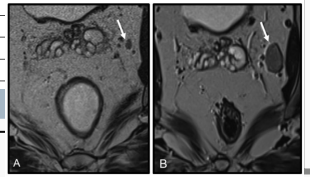

***Table 2*: Adaptive LASSO regression, SCAD, and MCP**

Panel B:

Non-Zero Coefficients for Path T Stage	*Adaptive Lasso*	*SCAD*	*MCP*
number_of_positive_lymph_n	**0.073**	**0.054**	
Sex		**-0.367**	**-0.340**
init_clinical_staging_m		**-0.174**	**-0.217**
Bmi		**0.158**	**0.058**
days_from_diagnosis_to_surgery		**-0.190**	**-0.132**
distance_to_distal_margin		**0.079**	**0.026**
number_of_lymph_nodes_exam		0.024	

Non-Zero Coefficients for Path N Stage	*Adaptive Lasso*	*SCAD*	*MCP*
number_of_positive_lymph_n	**0.214**	**0.144**	**0.060**
Race		**0.177**	**0.297**
init_clinical_staging_m		**0.217**	**0.346**

Non-Zero Coefficients for Path M Stage	*Adaptive Lasso*	*SCAD*	*MCP*
number_of_positive_lymph_n	0.044		
Race		**0.209**	**0.200**
init_clinical_staging_m		**0.148**	**0.206**
Bmi		-0.046	
days_from_neo_xrt_to_surgery		0.047	
distance_to_proximal_margin		-0.018	
distance_to_distal_margin		-0.044	
number_of_lymph_nodes_exam		**-0.035**	**-0.030**

Non-Zero Coefficients for Recurrence	*Adaptive Lasso*	*SCAD*	*MCP*
init_clinical_staging_m		-0.024	

***Table 3*: Ridge and ElasticNet regression**

Panel A:

Significant Coefficients for Path N Stage	*Ridge*	*ElasticNet*
number_of_positive_lymph_n	**0.691**	**0.656**
lymphovascular_invasion	0.147	

Panel B:

Significant Coefficients for Path N Stage	*Ridge*	*ElasticNet*
init_clinical_staging_m	0.177	
number_of_positive_lymph_n	**0.514**	**0.562**
large_vessel_invasion	-0.139	

Discussion and Conclusion

Results support my hypothesis that among the various pre-surgery and MRI variables available for colorectal cancer patients, number_of_positive_lymph_n (imaging variable), and init_clinical_staging_m and race (pre-surgery variable) appear to be significantly associated with pathologic TNM and recurrence. The number_of_positive_lymph_n refers to the number of lymph nodes to which cancer has spread, also known as the n-stage. init_clinical_staging_m refers to clinical M, or metastatic, stage, determined at diagnosis prior to any treatment. There are only 55 cases in this analysis, so this is a small sample, yet offers multidimensional data. As the number of cases grows, further testing can be done, and the results can be more reliable, for instance, with the inclusion of an initial fit period and a subsequent test period. With advancing imaging technologies, the role of imaging in predicting and improving patient outcomes is likely to grow even more significant in the future.