Replicating/Improving for Adjuvant Therapy in Stage III Colon Cancer

Shangrong Chi

Background/Introduction

Colon cancer is an important cause of mortality worldwide. The paper shows that adjuvant therapy improved survival in stage III colon cancer patients and compares the therapy with the observation only to levamisole alone or fluorouracil plus levamisole. This analysis aims to replicate and improve the paper study by utilizing the 'colon' dataset from the R survival package.

Study Aim

Determine a better model from 2 models – cox proportional hazards model and Penalized Regression to prove the effectiveness of adjuvant therapies and adjust to all confounders.

Study Hypotheses

- Primary Hypothesis: The combination of fluorouracil and levamisole or levamisole alone or observation only improves cure rates compared to levamisole alone.
- Secondary Hypothesis: Significant confounders change will optimize and refine model parameters.

Study Population

- Inclusion Criteria: Patients diagnosed with stage III colon cancer and had complete records.
- Exclusion Criteria: Patients with incomplete survival data or unknown treatment groups.

Data Acquisition

- Dataset: 'Colon' dataset from the survival package in R.
- Description: The dataset includes colon cancer patients' information on treatment experiments and survival outcomes.
- OS: etype = 2 is death and we can focus on this part.

Outcomes and Variables

Outcome	Description	Variables and Source	Specifications

Primary: Overall Survival (OS)	Time from surgery to death from any cause	Survival dataset	status = 1 means the event has occurred, status = 0 means the observation was censored, and etype = 2 is death.
Secondary: None			

Additional Variables:

id: Patient ID;

study: Study indicator; it's 1 for all patients in this dataset;

rx: Treatment arm: observation, levamisole, or fluorouracil + levamisole;

age: continuous variable (years);

sex: male or female, 1: Male;

obstruct: 1: tumor obstructed the colon;

nodes: number of positive lymph nodes;

extent: Extent of the tumor's local spread;

perfor: 1: Yes (perforation occurred);

adhere: 1: Yes (adherence observed);

time: Number of days until the event (recurrence or death) or censoring (end of study or loss to follow-up);

differ: Tumor differentiation level;

surg: 1: Long interval (late registration after surgery);

node4: 1: Yes ($\geqslant$ 4 positive lymph nodes).

Statistical Analysis Plan

Data Preparation

- Check dataset.
- Simply remove observations with missing dataset.
- Replace the nodes = 0 to the nodes = 1.
- Filter dataset with etype = 2.
- Summary(dataset) by treatment arm.

- RStudio: Notebook Output

A tibble: 20 × 4

Variable <chr>	Obs <chr>	Lev <chr>	Lev+5FU <chr>
Male_n	161	164	135
Male_Percent	52.79	55.78	46.71
Female_n	144	130	154
Female_Percent	47.21	44.22	53.29
Age_Mean_SD	59.47 (12.05)	60.09 (11.74)	59.9 (11.99)
Obstruct_Yes_Percent	20.33	19.73	17.65
Perfor_Yes_Percent	2.95	3.4	2.77
Adhere_Yes_Percent	14.75	15.31	13.15
Nodes_Mean_SD	3.84 (3.75)	3.71 (3.6)	3.44 (3.23)
Time_Mean_SD	1599.68 (855.81)	1625.09 (894.23)	1804.52 (858.38)
Differ_Well_Percent	8.85	12.59	9
Differ_Moderate_Percent	74.1	73.13	73.36
Differ_Poor_Percent	17.05	14.29	17.65
Extent_Submucosa_Percent	2.3	1.02	3.11
Extent_Muscle_Percent	11.8	11.9	10.73
Extent_Serosa_Percent	80.33	83.67	82.7
Extent_Contiguous_Percent	5.57	3.4	3.46
Surg_Short_Percent	70.49	74.15	75.09
Surg_Long_Percent	29.51	25.85	24.91
Node4_Yes_Percent	27.54	27.55	24.22

20 rows

Analyses Plan 1: Fit Models

- Kaplan-Meier Survival Analysis:

1. Plot Kaplan-Meier curves stratified by rx.
2. Perform log-rank tests to compare survival distributions.

- Cox Proportional Hazards Model:

1. Do not adjust and just apply treatments.
2. Adjust for all confounders such as age, sex, nodes and others and treatments.
3. Test proportional hazard assumptions using Schoenfeld residuals and check hazard ratios, p-values and 95% CI.

- LASSO Penalized Regression:

1. Use LASSO Penalized Regression to choose variables.

Analyses Plan 2: Adjust Confounders

- Check the changes for confounders using LASSO model, HR estimates and treatment effect.

Model Comparison and Final Model Selection

- Compare the cox model and LASSO model using:
    - AIC/ BIC criterion.
    - Likelihood ratio tests to compare models.

    - Harrell's C-index and other methods for discrimination.
    - Calibration using KM curves for risk groups and other plots.
    - Interpret.
- Select the final model based on better fit, discrimination, and calibration results.

Software

- Statistical Software: R version.
- Packages: survival, glmnet, ggplot2, dplyr, tidyr.