This dataset is from the National Cancer Data Base on patients that received adjuvant chemotherapy after non-small cell lung cancer (NSCLC) resection. Studies have reported that adjuvant chemotherapy offers a survival benefit to a number of staging scenarios in NSCLC. Variable recovery from lung cancer surgery may delay a patient’s ability to tolerate adjuvant chemotherapy, yet the urgency of chemotherapy initiation is unclear. The objective of this study is to determine the relationship between adjuvant treatment timing and efficacy. Students are expected to conduct statistical analysis to (1) investigate the relationship of survival and the time between NSCLC resection and the start of post-operative chemotherapy, and (2) identify the optimal treatment interval for the initiation of adjuvant chemotherapy following surgical resection.

The outcome variable is a binary indicator of death at 4 years, i.e.

* Cases (4318 patients): Vital\_status = 0 & Survival\_mo ≤ 48
* Controls (4396 patients): Survival\_mo > 48

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Value-Key** |
| Year\_diag | Year the patient was diagnosed with their cancer |  |
| Facility\_type | The classification of the facility the patient was treated at | 'Academic'  'Nonacademic'  'Unknown'; |
| Facility\_location | The geographic region that the patient’s treating facility was located in | 'North East'  'South Atlantic' 'Central'  'Mountain'  'Pacific'  'Other'; |
| Sex | Patient sex | Male  Female |
| Age | Patient age | Continuous |
| Insurance | The patient’s primary insurer | 'No Insurance'  'Private'  'Medicaid'  'Medicare'  'Other'  'Unknown'; |
| Income | The average income based on the patient’s zip code of residence | '<$48,000'  '$48,000+'  'Unknown'; |
| Education | The percent of people in the patient’s area of residence without a high school diploma | '21% or More'  '13-20.9%'  '7-12.9%'  '< 7%'  'Unknown'; |
| CD\_score | The patient’s score on the Charlson-Deyo comorbidity index, a measure of their medical comorbidities | '0'  '1'  '2+'; |
| Primary\_site | The primary site of the patients tumor | 'Upper Lobe'  'Middle Lobe'  'Lower Lobe'  'Overlapping Lesion'  'Lung, NOS'; |
| Histology | The histology of the patient’s tumor | 'Large Cell Carcinoma'  'Squamous Cell Carcinoma' 'Adenocarcinoma' 'Other'; |
| Grade | The histologic grade of the patient’s tumor | '1'  '2'  '3'  '4'  'Unknown'; |
| Path\_stage | The patient’s pathologic stage | 'I'  'II'  'III'; |
| Tumor\_size | The size of the tumor in millimeters |  |
| Surgery | Describes the patient’s treatment strategy | 'Lobectomy'  'Pneumonectomy'; |
| Chemo\_days | Start of post-operative chemotherapy after resection | Continuous |
| Vital\_status | Describes the patient’s vital status | 0 = 'Dead'  1 = 'Alive'; |
| Survival\_mo | Months after resection when vital status was recorded | Continuous |

Final report is suggested to follow the format of JAMA papers.

* **Materials and methods** section covers (1) info on variables (2) statistical analysis
* **Results** section covers (1) main results (2) tables and figures
* **Discussion** section covers (1) a general description of the study (2) conclusions (3) limitations

For students who use their own dataset, the report should also include the **Introduction** section to describe the aim of the study.

Don’t put all tables and figures from SAS output in the results section, only present relevant tables/figures. SAS code can be put in the Appendix. The final report should not exceed 3,000 words. Appendix does not count towards words limit.

All students must work independently and CANNOT discuss the final project with fellow students. If you have any questions, talk to the instructor before the last class on December 7th. After that, instructor will answer questions only by email. The due date is 5pm on Dec 16. Submit your final report electronically at Drop Box on Classes\*v2.