Using EHR Data to Identify Social Determinants of Health Affecting Disparities in Cancer Survival

WANTING CUI

JOSEPH FINKELSTEIN

Objectives

Introduction

Dataset

A limited dataset was generated from electronic medical records from Mount Sinai Health System

Time period: 2003/01/01 – 2019/12/31

Cancer types:

- Breast Cancer
- Colorectal Cancer
- Lung Cancer
- Multiple myeloma
- Prostate Cancer

Exclusion criteria:

- Children (less than 18 years old at diagnosis)
- Patients with missing values

Subsets and Variables

We extract 2 subsets:

- Short term survival: Patients who survived less than a year between diagnosis and death
- Long term survival: Patients who survived 5 year or more after diagnosis
  - Currently alive: patients' diagnoses date prior to 2015/1/1
  - Currently deceased: number of days between diagnoses and death are greater than 1825 days

Variables:

- Age
  - young adults: 18 – 40 years old
  - Middle age adults: 41 – 65 years old
  - Older adults: 66 and 66+ years old
- Sex, race, alive indicator, cancer diagnoses date and death date

Statistical Methods

Exploratory data analysis

Logistic Regression

- the effect of demographical factors on patients duration of survival after cancer diagnosis
- Independent variables: age, sex and race
- Dependent variables: short term survival (1), long term survival (0)

Results (EDA)

There are 22,096 patients in the dataset

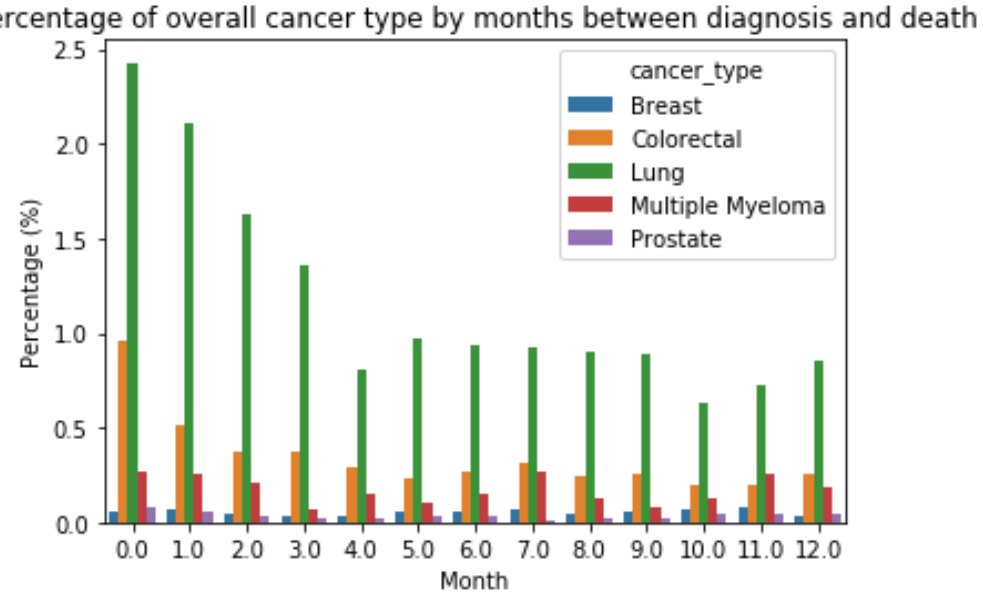
	Short Term Survival		Long Term Survival		Overall	
Cancer Type	Count	Percent	Count	Percent	Count	Percent
Breast	78	6.09%	6390	30.70%	6468	29.27%
Colorectal	292	22.81%	3613	17.36%	3905	17.67%
Lung	757	59.14%	2064	9.92%	2821	12.77%
Multiple Myeloma	92	7.19%	2017	9.69%	2109	9.54%
Prostate	61	4.77%	6732	32.34%	6793	30.74%
	1280		20816		22096	

Results (EDA)

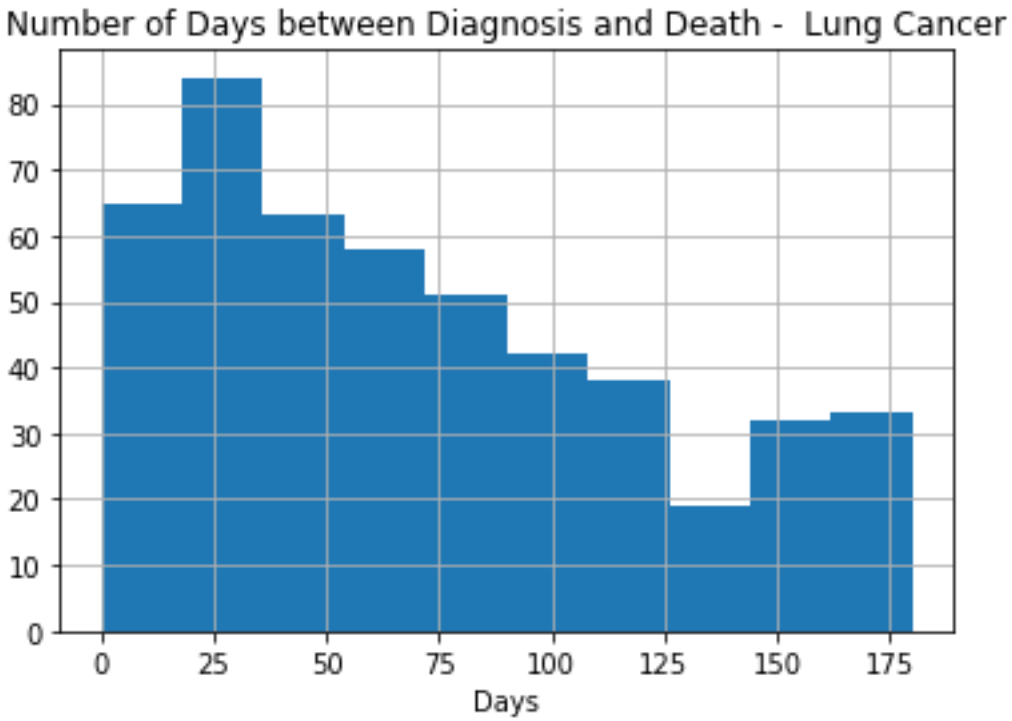
There are significantly more lung and colorectal cancer patients survived a short period of time

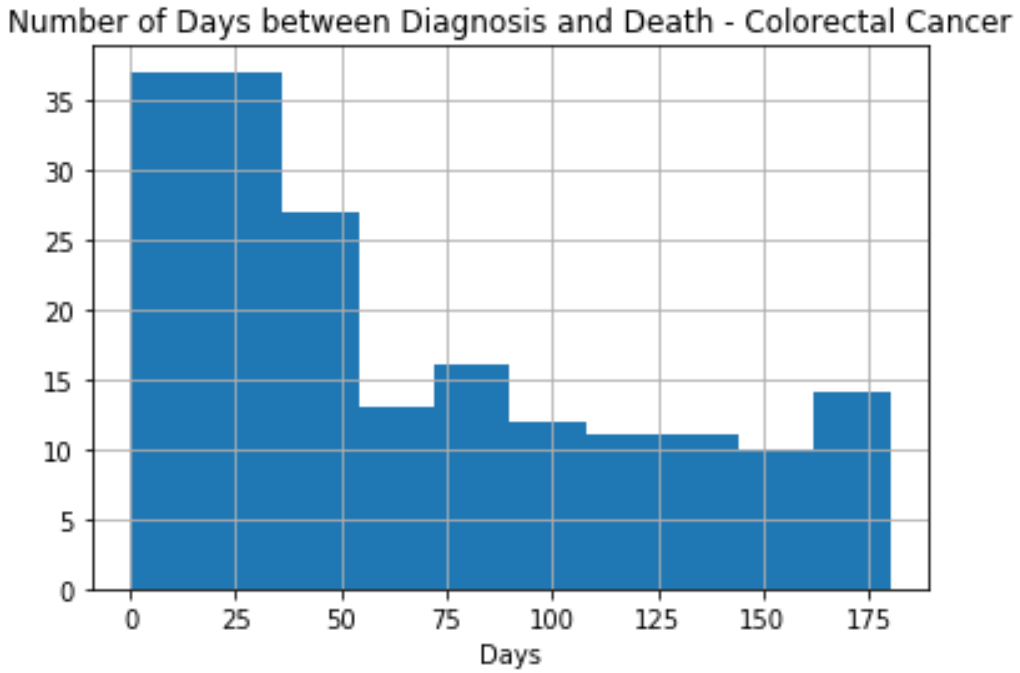
There are significantly less breast and prostate cancer patients survived a short period of time

The MM



Results (EDA)


Number of Days between Diagnosis and Death - Lung Cancer
0
10
20
30
40
50
60
70
80
0
25
50
75
100
125
150
175
Days


Number of Days between Diagnosis and Death - Colorectal Cancer
0
5
10
15
20
25
30
35
0
25
50
75
100
125
150
175
Days

Cancer Type	Lung		Colorectal		Breast		MM		Prostate	
Survival Period	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long
Count	757	2064	292	3613	78	6390	92	2017	61	6732
Age										
mean	69.84	67.23	71.74	62.41	71.38	56.38	67.93	59.38	73.48	62.27
std	11.16	11.05	15.04	13.84	15.71	12.98	12.34	10.95	10.62	8.40
median	70	68	74	62	70.5	56	67	60	73	62
Age group										
Young Adult	0.53%	1.50%	3.08%	5.90%	0.00%	10.03%	2.17%	4.71%	0.00%	0.30%
Middle Age Adult	33.29%	40.16%	27.40%	52.53%	37.18%	65.49%	42.39%	66.44%	22.95%	64.97%
Older Adult	66.18%	58.33%	69.52%	41.57%	62.82%	24.48%	55.43%	28.85%	77.05%	34.73%

Gender										
Female	44.65%	57.22%	44.52%	49.82%	100.00%	99.33%	40.22%	45.51%	0.00%	0.00%
Male	55.35%	42.78%	55.48%	50.18%	0.00%	0.67%	59.78%	54.49%	100.00%	100.00%
Race										
American Indian	0.26%	0.10%	0.00%	0.08%	0.00%	0.09%	0.00%	0.10%	0.00%	0.07%
Asian	0.66%	1.70%	1.71%	2.82%	0.00%	3.97%	0.00%	1.54%	0.00%	0.86%
Black	23.25%	12.06%	17.47%	13.01%	34.62%	13.97%	27.17%	16.86%	24.59%	15.21%
Islander	4.49%	3.68%	8.22%	7.61%	1.28%	1.61%	3.26%	1.04%	0.00%	0.94%
Other	16.38%	11.68%	14.73%	14.89%	28.21%	20.22%	20.65%	27.17%	16.39%	15.08%
White	54.95%	70.78%	57.88%	61.58%	35.90%	60.13%	48.91%	53.30%	59.02%	67.84%