Equity in Care:
The effect of socio-economic factors in late cancer diagnosis

Parviz Asoodehfard
Elba Gomez
Allen Lee
Tudor Schlanger

Motivation

What we know:

Early diagnosis and treatment are crucial for survival of cancer patients

What we may have:

Barriers to health access and education may still exist for certain groups

What we want:

Understanding how demographics affect early diagnosis can help create targeted Public Service Announcements to spread awareness

Research Questions

Question 1: What is the socio-economic profile of the typical individual who is diagnosed late?

Question 2: What is the impact of socio-economic traits/determinants on late diagnosis?

Question 3: What is the effect of late diagnosis and socio-economic traits on the survival interval after diagnosis?

Assumptions & Data Cleaning

Sample selection

Restricted to 2010-2014 when cancer diagnostic available

Restricted to patients where stage of cancer identified

Restricted to observations where patients confirmed deceased before 2014

Assumptions & Data Cleaning

Sample selection

Restricted to 2010-2014 when cancer diagnostic available

Restricted to patients where stage of cancer identified

Restricted to observations where patients confirmed deceased before 2014

Resulting sample size: 13,678

Assumptions & Data Cleaning

Sample selection

Restricted to 2010-2014 when cancer diagnostic available

Restricted to patients where stage of cancer identified

Restricted to observations where patients confirmed deceased before 2014

- **Data assumptions**

Census 2006 variables are time-invariant (they stay constant for our sample from 2010-2014)

Census variables are an accurate representation of socio-economic status

Q1.

What is the socio-economic profile of the typical individual who is diagnosed late?

In other words...
Who is the average Joe who gets diagnosed late?

Clustering

1. For $k = 2, \ldots, 10$

    i. Train and fit a *k*-means model on the socio-economic data

    ii. Compute proportion of late stage diagnoses for each cluster

2. Choose *k* to maximize the variance of the calculated proportions between different groups

3. Identify cluster with worst proportion of late stage diagnosis

4. Use cluster centroid as representative for the group

Typical Profile Late Diagnosis Patient

- Male
- Non-college educated
- Non-immigrant, non-visible minority
- Age of diagnosis: 58
- Married
- Family size: 2.5
- Low-income area-based deciles: 6
- Region: Ontario

Measuring Impact of Policy Action

Average number of days survived with cancer:

- **Diagnosed early:** 491
- **Diagnosed late:** 270

Q2.

What is the impact of socio-economic traits on late diagnosis?

In other words…
What factors matter in achieving early diagnosis?

Logistic Regression: All Cancers

0 = (Diagnosis stage 0, 1, 2) and 1 = (Diagnosis stage 3,4)

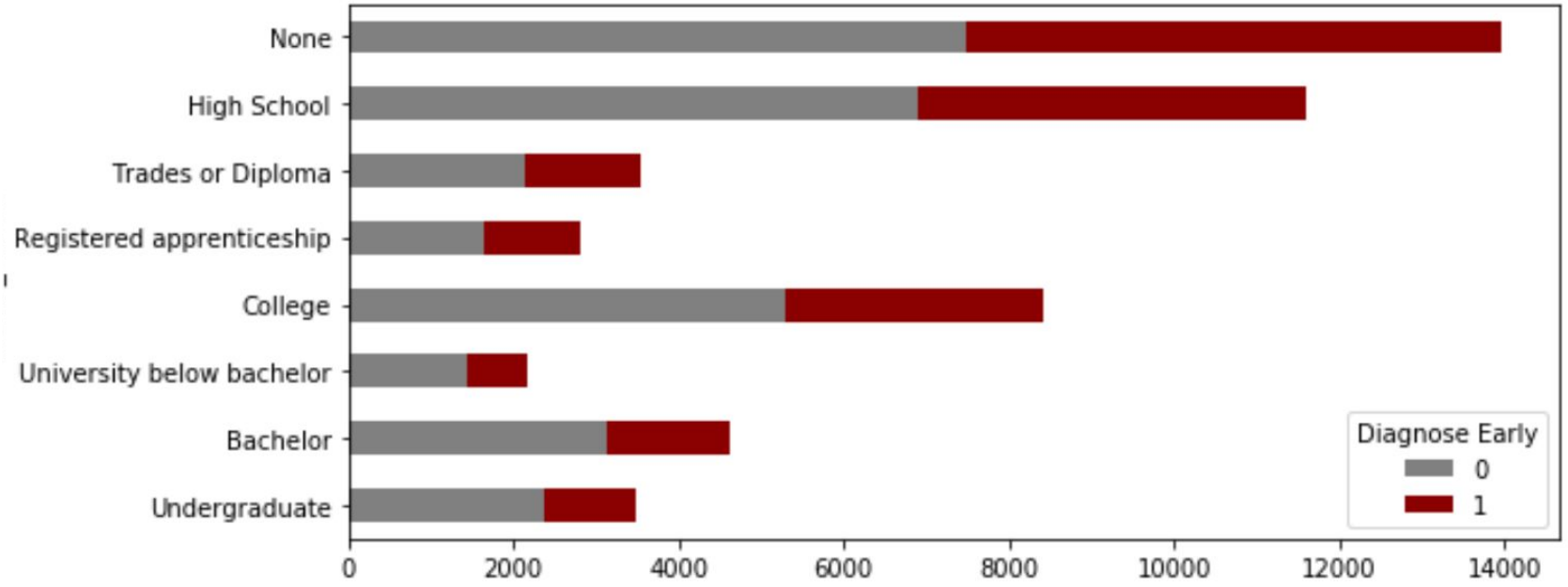
	coef	std err	z	P>|z|
Age at Diag	-0.0223	0.002	-10.917	0.000
Educational Attainment	0.0088	0.051	0.171	0.864
Immig. Status	-0.0337	0.051	-0.659	0.510
Married	-0.0216	0.075	-0.289	0.773
Separated	0.0882	0.143	0.618	0.537
Single	0.1175	0.105	1.117	0.264
Widowed	-0.0293	0.089	-0.330	0.741
decile_income_High	0.0950	0.047	2.040	0.041
Male	-0.0991	0.055	-1.815	0.070

Logistic Regression: Prostate Cancer

0 = (Diagnosis stage 0, 1, 2) and 1 = (Diagnosis stage 3,4)

	coef	std err	z	P>|z|
Age at Diag	0.0045	0.006	0.749	0.454
Educational Attainment	-0.3874	0.135	-2.872	0.004
Immig. Status	0.1584	0.135	1.173	0.241
Married	-0.2125	0.216	-0.984	0.325
Separated	0.6081	0.436	1.396	0.163
Single	-0.1747	0.317	-0.551	0.582
Widowed	-0.5094	0.283	-1.800	0.072
decile_income_High	0.1382	0.124	1.113	0.266
Male	-0.1349	0.484	-0.279	0.780

Educational Attainment for Cancer Population



Q3.

What is the effect of late diagnosis and socio- economic traits on the survival interval after diagnosis?

In other words...
Who lives longer and why?

Late diagnosis

Univariate regression

	coef	std err	t	P>|t|
const	491.6819	6.160	79.821	0.000
Late	-221.0619	6.974	-31.700	0.000

Multivariate regression

	coef	std err	t	P>|t|
const	155.7146	7.088	21.967	0.000
Late	-154.7483	7.132	-21.697	0.000
Size Family	3.1179	2.734	1.140	0.254
Age Diagnostic	-2.0956	0.271	-7.747	0.000

Relevant socio-demographic factors

	coef	std err	t	P>|t|
College Educated_2	13.7138	6.513	2.106	0.035
Immigrant Dummy_2	26.8721	6.616	4.062	0.000
Marital Status_2.0	3.6338	9.563	0.380	0.704
Marital Status_3.0	-27.2160	16.688	-1.631	0.103
Marital Status_4.0	-20.9218	12.555	-1.666	0.096
Marital Status_5.0	1.4858	11.192	0.133	0.894
Urban Dummy_2.0	0.5913	6.425	0.092	0.927
Income Decile_2	17.4923	5.898	2.966	0.003
Male_2.0	-16.2747	6.381	-2.551	0.011

What is not so important for survival

- Province of diagnosis
- Native status and most visible minorities
- Marital status
- Rural/urban divide

Conclusion

Typical profile:

But **education** is a key factor

Limitations

- Different types of cancers have different characteristics
  - Should do separate analysis for each type of cancer

- Using only demographic data from the 2006 census
  - We expect certain demographic and economic data to change throughout time

- Results may be biased because certain population groups (e.g. Native people) are not being diagnosed at all, so samples are not representative

- Controls for other risky behaviour not included: smoking, alcohol and drug use

Next Steps: Additional Data

- **Data**
  - Look at late diagnosis by geographic area (postal-code level)
  - Look at immigration status in a more careful way (Longitudinal Immigration Database)
  - With a additional census data (e.g. 2016 census), we can get a more holistic view of each patient

- **Economic Data (TIFF)**
  - Employment Insurance and Social Assistance data only available for employed population; how does it affect early diagnosis?
    - Can tap on financial burden of care and importance of out of pocket costs

Thank you very much!