PREDICTING THE POPULARITY OF TED TALKS USING COMPOSITE MEASURES OF POPULARITY AND MIXED MODELS

Marcos Cardenas Zelaya

March 10, 2020

ABSTRACT

Objectives: Determine the relationship between osteoarthritis (OA) and cardiovascular disease using Canadian survey data.

Design: Logistic Mixed-Models Regression is used to determine the odds ratio between OA and heart disease.

Data: Canadian Community Health Survey (CCHS) from 2000 to 2005.

Participants: Adult participants aged 20-64 in the CCHS cycles 1.1, 2.1 and 3.1 were included. CCHS dataset includes nationally representative data on heart disease and other health determinants. All observations (responses and predictors) are self-reported from 10 provinces and 3 territories. We have selected 200,478 observations. Observations are not identifiable between cycles.

Predictors and Response: Cardiovascular disease is the response. The main predictor is OA after adjusting for socio-demographic factors, access to a doctor, obesity, physical activity, smoking status, drinking status, diabetes and hypertension.

Results: There is no evidence to suggest that OA is associated with heart disease. There is also little evidence to suggest that the association between OA and heart disease vary across gender, marital status, region and recency of immigration.

Conclusion: Accounting for demographics, OA is not associated with heart disease. Due to computational restrictions, more research is required to accurately asses the relationship between OA and heart disease.

1 Introduction

2 Methods

2.1 Data Sources

2.2 Study Sample

3 Response and Predictors

3.1 Response: Heart Disease

3.2 Predictors

3.2.1 Exposure/Risk Factor: Osteoarthritis

3.3 Covariates & Confounders

4 Statistical Analysis

4.1 Logistic Regression Cross-Section

4.2 Logistic Mixed-Models Cross-Section

4.2.1 Regional Variation

Variable	Levels	n	%	$\sum\%$
title_sentiment_tfidf	future	2197	86.2	86.2
	life	85	3.3	89.5
	new	73	2.9	92.3
	world	195	7.7	100.0
	all	2550	100.0	
tags_label_tfidf	brain	147	5.8	5.8
	business	184	7.2	13.0
	culture	605	23.7	36.7
	design	329	12.9	49.6
	energy	64	2.5	52.1
	global	354	13.9	66.0
	health	192	7.5	73.5
	music	118	4.6	78.2
	science	346	13.6	91.7
	social	211	8.3	100.0
	all	2550	100.0	

Table 1

Variable	n	Min	q_1	$\widetilde{x}$	$\bar{x}$	q_3	Max	s	IQR	#NA
avg_views_per_day	2550	17.0	311.0	724.0	1486.1	1752.8	28347.0	2148.2	1441.8	0
duration_no_norm	2550	135.0	577.0	848.0	826.5	1046.8	5256.0	374.0	469.8	0
num_speaker_no_norm	2550	1.0	1.0	1.0	1.0	1.0	5.0	0.2	0.0	0

film_age_no_norm	2550	126.0	1177.2	2100.0	2230.9	2977.0	16667.0	1385.9	1799.8	0
title_length_no_norm	2550	1.0	5.0	6.0	6.2	8.0	16.0	2.3	3.0	0
popularity	2550	0.6	1.6	1.8	1.7	1.9	3.2	0.2	0.2	0

Table 2

5 Results

5.1 Poisson & Linear Regression

lasso was used for model selection

5.2 Poisson & Linear Mixed-Effects

5.2.1 Model Selection

Table 3: Testing Random Slopes

	Df	Chisq	Chi Df	Pr($>$Chisq)
normal_popularity_themes_1	10			
normal_popularity_themes_2	45	59.91	35	0.0055
normal_popularity_times_1	10			
normal_popularity_times_2	19	46.89	9	0.0000
pois_views_themes_1	9			
pois_views_themes_2	44	111138.01	35	0.0000
pois_views_times_1	9			
pois_views_times_2	44	111138.01	35	0.0000

Table 4: Testing Random Intercept

	df	AIC
normal_popularity_themes_0	9.00	-810.37
normal_popularity_themes_1	10.00	-789.73
normal_popularity_times_0	9.00	-810.37
normal_popularity_times_1	10.00	-789.73
pois_views_themes_0	8.00	2456258.87
pois_views_themes_1	9.00	2262017.56
pois_views_times_0	8.00	2456258.87
pois_views_times_1	9.00	2262017.56

6 Conclusion

References