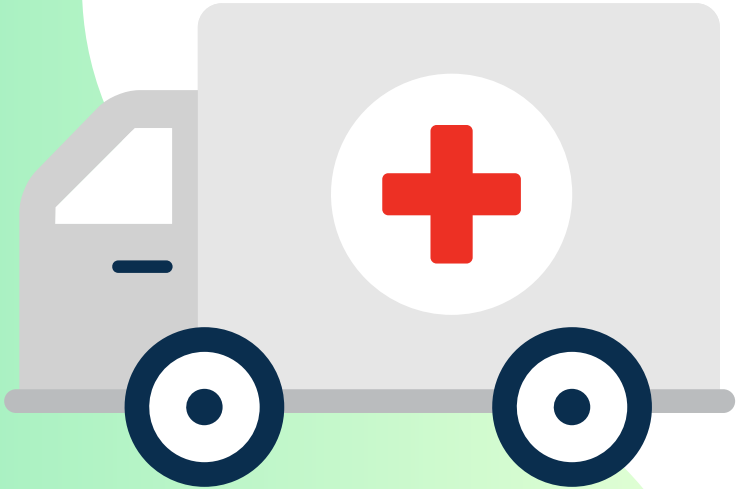
The Causal Effects of pH1N1 Vaccination

Rachel Flodin, Kaelan Yu, and Natalie Chisam

Table of Contents

575,400+

worldwide died from (H1N1)pdm09 virus infection in its first year in circulation



Understanding likelihood to receive a vaccine is more relevant than ever amidst the COVID-19 global pandemic.

Project Overview

Investigate the causal effects of a patient's likelihood to receive the new pH1N1 vaccine.

Research Questions

Does receiving the seasonal flu vaccine have an effect on receiving the pH1N1 vaccine?

Does receiving a doctor recommendation have an effect on receiving the pH1N1 vaccine?

Data

Source

National 2009 H1N1 Flu Survey (NHFS)

Sponsored

National Center for Immunization and Respiratory Diseases (NCIRD)

Administered

NCIRD and the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC)

Sample

26,707 US-based participants selected random-digit-dialing telephone

Time

October 2009 - June 2010

Variables

38 variables, including questions regarding vaccination and personal data

Data Preprocessing

1. Initial Variable Selection

38 variables → 34 variables; removed:

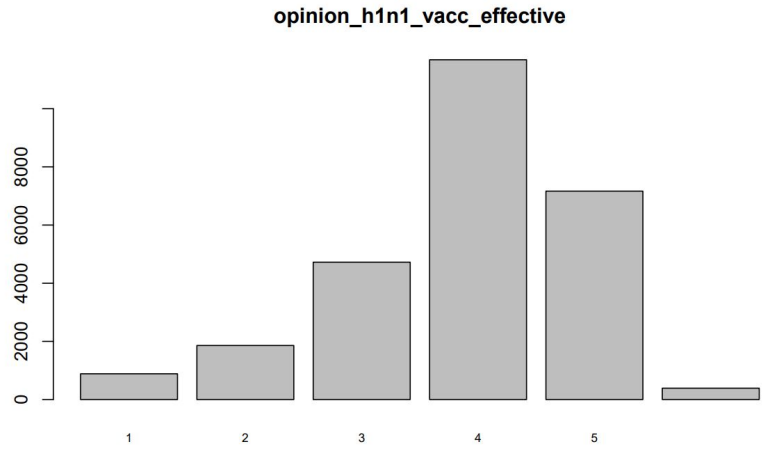
- respondent_id
- hhs_geo_region
- employment_industry
- employment_occupation

2. Missing (NA) Values (55.8%)

- initial sample size: n = 26707
- final sample size: n = 11794
- we are still in n >> p setting! :)

3. Variable Types

- all categorical so we visualize with bar plots
  - nominal (unordered)
  - ordinal (ordered)



Exploratory Data Analysis

1. Model Fitting (logistic regression)

`glm()` with logit link, binomial family

- response: h1n1_vaccine

(whether respondent received vaccine)

2. Model Selection (BIC)

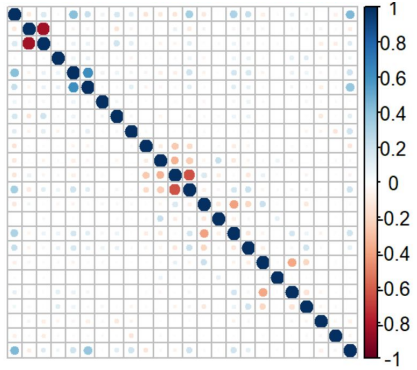
- BIC since large n (penalty $\propto$ log(n))
- best subset selection
- 34 $\rightarrow$ 13 variables

3. Model Diagnostics (check for multicollinearity)

(1) Variance inflation factors (VIFS)

- VIFs all below 5 so no significant multicollinearity ($r_i < 0.8$)

(2) Pairwise Correlation Plot



Proposed Analysis

Exploratory data analysis

- Model fitting, selection, and diagnostics

Causal inference

- Structure learning
- Methods for adjustment

Questions?