The Effect of Vaccination on Hospital Admission

https://github.com/erickim/PH252D_final_project

Denys Dukhovnov, James Duncan, Eric Kim, David Proudman

April 30, 2018

Introduction

- Americans ≥65 have some of the highest immunization rates among other age groups in the country
- Seasonal Influenza vaccine is the most common type with 63.4% of adults over 65 getting the vaccine (2015-2016)[1]
- The elderly subpopulation with type 2 diabetes must take additional precautions to due to comorbidities, such as various heart and respiratory tract conditions, among other

[1] Centers for Disease Control and Prevention. Recommended Vaccines for Adults. https://www.cdc.gov/vaccines/adults/rec-vac/index.html. Published January 25, 2018. Accessed April 17, 2018

Introduction (cont.)

- By 2015 the overall hospitalization rates among ≥65 have decreased to 26,400 from 35,214 per 100,000 in 2000 (25% decrease)[2]
- Hospitalization rates for cardiac and stroke, cerebrovascular disease, pneumonia/influenza, and other cause-of-death related conditions in the unvaccinated population ≥65 are close to 3% during the influenza season, but get significantly reduced to 2.2% for the seasonal influenza vaccine recipients[3]

---

[2] Sun R, Karaca Z, Wong HS. Trends in Hospital Inpatient Stays by Age and Payer, 2000-2015. Rockville, MD: Agency for Healthcare Research and Quality; 2018. https://www.hcup-us.ahrq.gov/reports/statbriefs/sb235-Inpatient-Stays-Age-Payer-Trends.jsp. Accessed April 23, 2018.

[3] Nichol KL et al. Influenza Vaccination and Reduction in Hospitalizations for Cardiac Disease and Stroke among the Elderly. *N Engl J Med*. 2003;348(14):1322-1332.

Scientific Question

What is the effect of [any] vaccination in 2015 on hospital admission [for any reason] in 2016 among the older adults over 65 years old, diagnosed with diabetes?

The Causal Model and Target Causal Parameter

Endogenous variables:

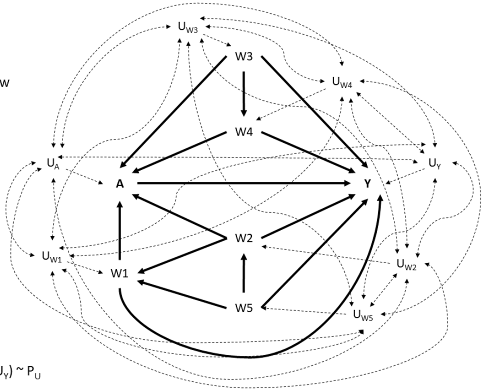
A: Vaccination in 2015
Y: Hospital admission in 2016
W1: Income + Net worth
W2: College education above/below
W3: Age
W4: Medical risk score
W5: Sex

- No independence assumptions
- Exclusion restriction: W5 on A

Structural Equations:

$A = f_A(W1, W2, W3, W4, U_A)$
$Y = f_Y(W1, W2, W3, W4, W5, U_Y)$
$W1 = f_{W1}(W2, W5, U_{W1})$
$W2 = f_{W2}(W5, U_{W2})$
$W3 = f_{W3}(U_{W3})$
$W4 = f_{W4}(W3, U_{W4})$
$W5 = f_{W5}(U_{W5})$

$U = (U_{W1}, U_{W2}, U_{W3}, U_{W4}, U_{W5}, U_A, U_Y) \sim P_U$



Our target causal parameter is the average treatment effect

$$\Psi^F(P_{U,X}) = \mathbb{E}_{U,X}[Y_1 - Y_0]$$

Where the counterfactual variables of interest is the number of hospital admissions under vaccination and in the absence of vaccinations.

Observed Data

		Vaccinated		Gender		College		
		N	Y	F	M	N	Y	Overall
	count	20904	5875	14055	12724	24843	1936	26779
	% in stratum	78.06%	21.94%	52.48%	47.51%	92.77%	7.23%	100%
Wealth Index	mean	11.73	11.94	11.54	12.03	11.75	12.10	11.78
	sd	1.51	1.53	1.57	1.41	1.51	1.48	1.51
Age	mean	74.66	74.68	74.87	74.43	74.75	73.51	74.66
	sd	6.61	6.32	6.74	6.32	6.57	6.11	6.55
Medical Risk Score	mean	1.20	1.11	1.17	1.19	1.18	1.10	1.18
	sd	0.98	0.87	0.94	0.98	0.96	0.96	0.96
No. Hospital Visits	mean	0.34	0.29	0.32	0.33	0.33	0.28	0.33
	sd	0.94	0.83	0.90	0.93	0.92	0.86	0.92

Observed Data (cont.)

		Vaccinated							
		No				Yes			
		Gender				Gender			
		F		M		F		M	
		College		College		College		College	
		N	Y	N	Y	N	Y	N	Y
	count	10133	763	9308	700	2901	258	2501	215
	% of total	37.84%	2.85%	34.76%	2.61%	10.83%	0.96%	9.34%	0.80%
Wealth Index	mean	11.47	11.79	11.97	12.27	11.70	12.18	12.15	12.55
	sd	1.56	1.52	1.39	1.44	1.58	1.43	1.44	1.34
Age	mean	75.01	73.18	74.45	73.82	74.95	73.42	74.58	73.80
	sd	6.83	6.05	6.40	6.28	6.60	5.91	6.04	6.01
Medical Risk Score	mean	1.20	1.12	1.20	1.10	1.09	1.00	1.15	1.16
	sd	0.96	0.97	1.00	1.01	0.86	0.81	0.90	0.92
No. Hospital Visits	mean	0.34	0.30	0.34	0.30	0.28	0.17	0.31	0.25
	sd	0.93	0.95	0.95	0.92	0.82	0.50	0.88	0.66

Identifiability

Under the current model the target causal parameter is not identifiable. But given background knowledge, some independence assumptions can be established.

$U_A \perp\!\!\!\perp U_{W_3}, U_{W_5}$
$U_Y \perp\!\!\!\perp U_{W_1}, U_{W_2}, U_{W_3}, U_{W_5}$
$U_{W_1} \perp\!\!\!\perp U_{W_3}, U_{W_4}, U_{W_5}, U_Y$
$U_{W_2} \perp\!\!\!\perp U_{W_3}, U_{W_4}, U_{W_5}, U_Y$
$U_{W_3} \perp\!\!\!\perp U_A, U_{W_1}, U_{W_2}, U_{W_4}, U_{W_5}, U_Y$
$U_{W_4} \perp\!\!\!\perp U_{W_1}, U_{W_2}, U_{W_3}, U_{W_5}$
$U_{W_5} \perp\!\!\!\perp U_A, U_{W_1}, U_{W_2}, U_{W_3}, U_{W_4}, U_Y$

Knowledge is insufficient. Further convenience assumptions are required for identifiability, under the backdoor criterion and randomization assumptions.

$U_A \perp\!\!\!\perp U_Y, U_{W_3}, U_{W_4}, U_{W_5}$
$U_Y \perp\!\!\!\perp U_A, U_{W_3}, U_{W_5}$

Identifiability (cont.)

Endogenous variables:

A: Vaccination in 2015

Y: Hospital admission in 2016

W1: Income + Net worth

W2: College education above/below

W3: Age

W4: Medical risk score

W5: Sex

Convenience assumptions:

$U_A \perp\!\!\!\perp U_Y, U_{W3}, U_{W4}, U_{W5}$

$U_Y \perp\!\!\!\perp U_A, U_{W1}, U_{W2}, U_{W3}, U_{W5}$

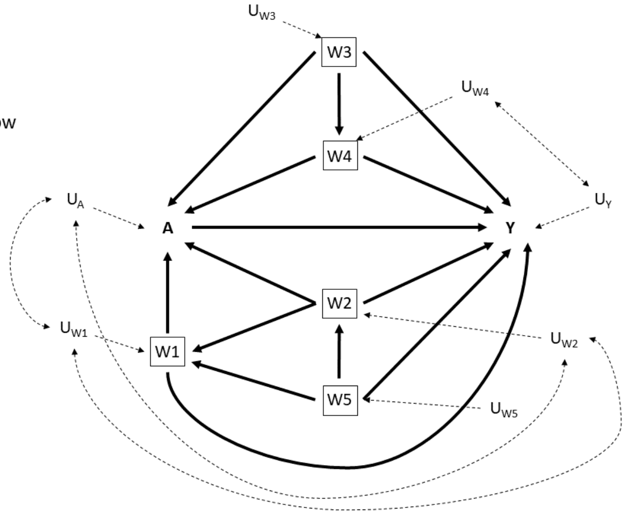
$U_{W1} \perp\!\!\!\perp U_{W3}, U_{W4}, U_{W5}, U_Y$

$U_{W2} \perp\!\!\!\perp U_{W3}, U_{W4}, U_{W5}, U_Y$

$U_{W3} \perp\!\!\!\perp U_A, U_{W1}, U_{W2}, U_{W4}, U_{W5}, U_Y$

$U_{W4} \perp\!\!\!\perp U_{W1}, U_{W2}, U_{W3}, U_{W5}$

$U_{W5} \perp\!\!\!\perp U_A, U_{W1}, U_{W2}, U_{W3}, U_{W4}, U_Y$



Statistical Model and Estimand

Our statistical model is a semi-parametric model where we take the previous assumptions on identifiability as restrictions on our observed data distributions P_0.

Our estimands for the Average Treatment Effect will be the G-Computation, IPTW, Stabilized IPTW, and TMLE estimators.

Math Review

G-computation Estimator

The G-computation/simple substitution estimator is as follows

$$\hat{\Psi}(\mathbb{P}_n) = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\mathbb{E}}\Big(Y_i \,|\, A_i = 1, \overrightarrow{W}_i\Big) - \hat{\mathbb{E}}\Big(Y_i \,|\, A_i = 0, \overrightarrow{W}_i\Big)\right]$$

IPTW Estimator

The IPTW substitution estimator is as follows (similar to Horvitz-Thompson estimator)

$$\hat{\Psi}(\mathbb{P}_n) = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}\{A_i = 1\}}{g(A_i \,|\, W_i)}Y_i - \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}\{A_i = 0\}}{g(A_i \,|\, W_i)}Y_i$$

More Math Review

Stabilized IPTW Estimator

The Stabilized IPTW substitution estimator is as follows (similar to Hajek estimator

$$\hat{\Psi}(\mathbb{P}_n) = \frac{\sum_{i=1}^{n} \frac{\mathbb{1}\{A_i=1\}}{g(A_i \mid W_i)} Y_i}{\sum_{j=1}^{n} \frac{\mathbb{1}\{A_j=1\}}{g(A_j \mid W_j)}} - \frac{\sum_{i=1}^{n} \frac{\mathbb{1}\{A_i=0\}}{g(A_i \mid W_i)} Y_i}{\sum_{j=1}^{n} \frac{\mathbb{1}\{A_j=0\}}{g(A_j \mid W_j)}}$$

Targeted Maximum Likelihood Estimator

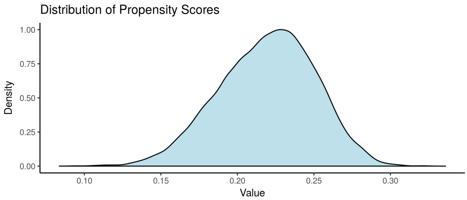
The TMLE estimator is defined to be the solution to the efficient influence curve

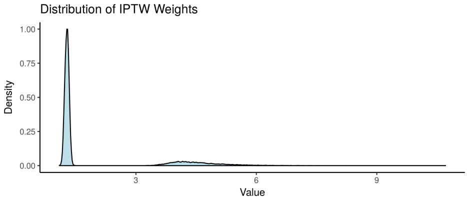
$$0 = \frac{1}{n} \sum_{i=1}^{n} D^*(P_n^*)(O_i)$$

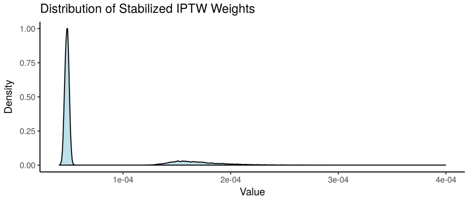
Where D^* is the efficient influence curve as follows

$$D^*(P) = \left[\frac{A}{g(A \mid W)} - \frac{1-A}{g(0 \mid W)} \right] [Y - \overline{Q}(A, W)] + \overline{Q}(1, W) - \overline{Q}(0, W) - \psi$$

Positivity Assumption





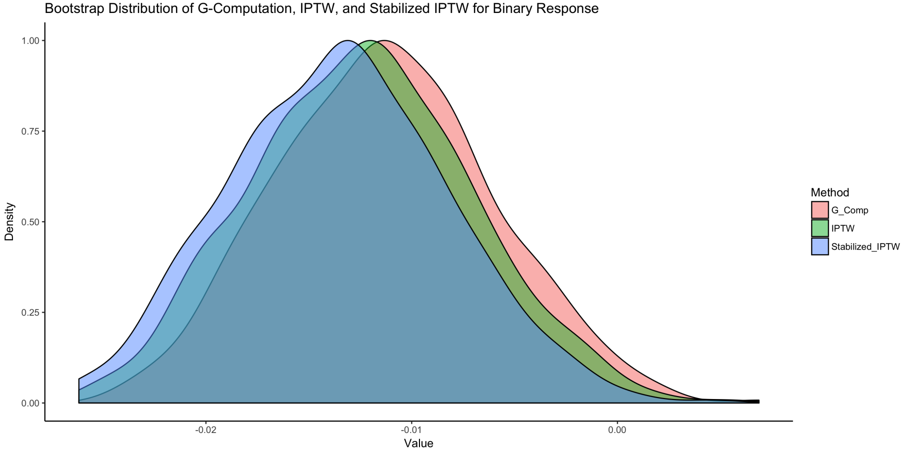


Propensity Scores	min	0.084
	median	0.221
	mean	0.219
	max	0.336
IPTW Weights	min	1.09
	median	1.30
	mean	2.00
	max	10.7
Stabilized IPTW	min	$4.08e-05$
	median	$4.86e-05$
	mean	$7.47e-05$
	max	$4.00e-04$

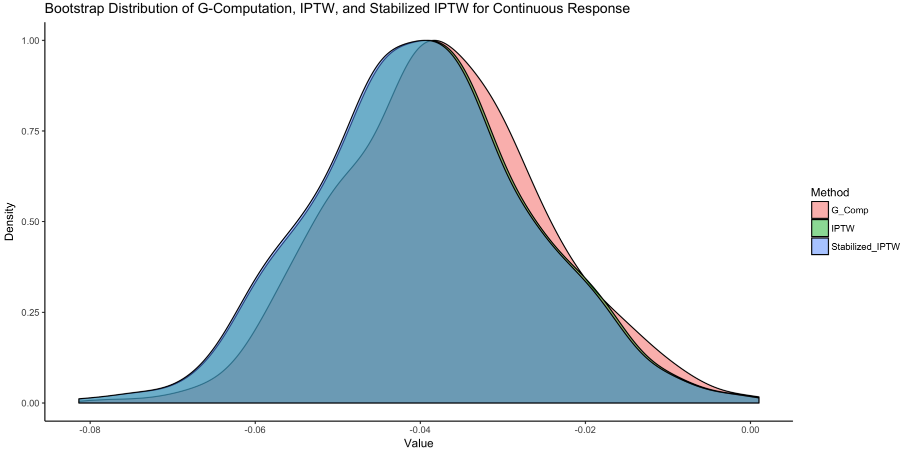
Results

	Outcome			
	Binary		Continuous	
Estimand	$\hat{\Psi}$	SE	$\hat{\Psi}$	SE
G-Computation	-0.0116	0.00525	-0.0373	0.0125
IPTW	-0.0127	0.00534	-0.0394	0.0130
Stabilized IPTW	-0.0139	0.00531	-0.0397	0.0130
TMLE (SuperLearner)	-0.0107	0.00550	-0.0318	0.0127

Inference: Bootstrap Distributions Binary Response



Inference: Bootstrap Distributions Continuous Response



Inference: Hypothesis Tests

Based on our estimates and the estimated standard error, we can test whether the treatment of vaccination has a significant effect on hospitalization. Testing a two sided hypothesis, we get the following *p*-values

	Outcome	
	Binary	Continuous
Estimand	*p*-value	*p*-value
G-Computation	0.0271	0.00285
IPTW	0.0174	0.00244
Stabilized IPTW	0.00885	0.00226
TMLE	0.052	0.0123

Conclusion

In all cases, we reject the null of no effect with the treatment. Interestingly, TMLE with SuperLearner had lower significance levels, but we note that because of our small estimates and standard errors, the p-values were rather sensitive to small changes.

For binary response, we conclude that there is around a 1% to 1.5% decrease in hospitalizations for the treatment of vaccinations, whereas for continuous response, we conclude that for every 100 people who get vaccinated, there is a decrease of around 3-4 hospital visits.