Assignment 1 Routine Care

Martijn Koster, 6234119   Jurrian van de Kraats, 5961688   Tim Poorthuis, 0651478

2023-02-15

1 Introduction

In this paper we will examine whether there is a causal effect from elderly people who have received an influenza vaccine and the likelihood of hospitalization. This will be examined by using the patient records dataset and different modeling methods using PS, to account for confounders.

2 Methods

2.1 Examening the causal structure

Domain knowledge was used to create a DAG (Figure 1) that explains the causal structure present in the data. The variable 'contact with chiropractor (GP)' forms a proxy for the unobserved variable healthy lifestyle. Getting the influenza vaccine is connected to age and contact with GP. Having obtained influenza vaccination was associated with lower risk for adverse cardiovascular events [Behrouzi et al., 2022]. Having pulmonary disease and diabetes increases the chance for cardiovascular disease (cvd) [Genootschap]. Sex forms a confounder between hospitalization and healthy lifestyle [Loef and Walach, 2012] which in turn causes the aforementioned diseases.

The causal paths of the DAG are examined to assess whether unbiased causal inference is possible. There is a flow of statistical information through open backdoor paths due to observed and unobserved confounders. The model would need to adjust for these confounders in order to perform unbiased causal inference.

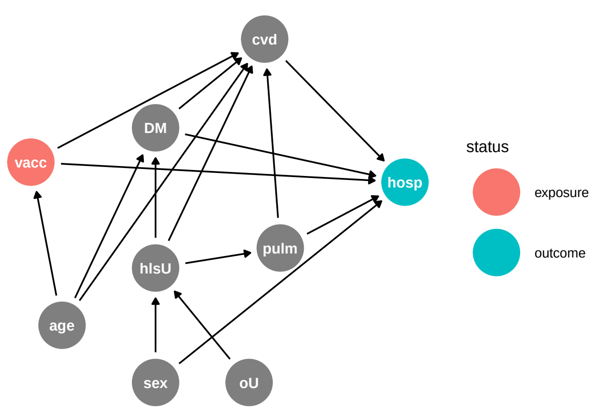

Figure 1: DAG model with Vaccine as exposure and hospitalization as outcome

2.2 Statistical Methods

2.2.1 Baseline Characteristics

The electronic patient records dataset consist of eight variables. The variables age and contact with a chiropractor were continuous, the variables vaccination status, sex, cvd, pulmonary disease, diabetes and hospitalization were binominal. Table 1 presents the baseline characteristics of 40000 individuals who were included in the study.

Baseline characteristics stratified by the study outcome indicate that 254 of the respondents were hospitalized. Respondents who were hospitalized were older (79.68 vs 75.63), had more contact with the GP (21.08 vs 14.71), were less often female (48.43%), had more often cvd (72.05% vs 49.25%) and more often Diabetes (11.42% vs 6.51%).

Table 1: Baseline Characteristics stratified by Influenza vaccination received and hospitalization

		Influenza vaccination received		Hospitalized	
Characteristics	Total	Yes	No	Yes	No
N	40000	29616	10384	254	39746
Age, mean (SD)	75.65 (6.97)	75.9 (6.83)	74.97 (7.32)	79.68 (7.2)	75.63 (6.97)
Contact, mean (SD)	14.75 (11.54)	15.85 (11.73)	11.64 (10.38)	21.08 (15.59)	14.71 (11.5)
Female, n (%)	24763 (61.91)	18022 (60.85)	6741 (64.92)	123 (48.43)	24640 (61.99)
Pulmonary disease, n (%)	4937 (12.34)	4244 (14.33)	693 (6.67)	60 (14.33)	4877 (12.27)
cvd, n (%)	19757 (49.39)	15702 (53.02)	4055 (39.05)	183 (72.05)	19574 (49.25)
Diabetes mellitus, n(%)	2618 (6.54)	2223 (7.51)	395 (3.8)	29 (11.42)	2589 (6.51)
Received Influenza Vaccination, n (%)				184 (72.44)	29432 (74.05)

2.2.2 Propensity Scores (PS)

A logistic regression was used to estimate the PS. A PS gives the probability being vaccinated for the respondents. Based on the DAG (Figure 1) the variables: age, sex, cvd, pulmonary disease, diabetes, and GP contact were used in the PS model. For the variables age and contact a spline is used [Tian et al., 2019].

In Figure 2, the PS for vaccinated and unvaccinated individuals appear to be well-balanced, supporting the positivity assumption, which means that both treatment groups have a chance to get the treatment given the covariates [Westreich and Cole, 2010].

2.2.3 PS as Covariate

The first adjusted model is to use the PS as a covariate. In a observational study the treated and untreated group have an equal distribution given that these are divided in groups of a constant propensity. The PS can be used as a baseline variable to account for the dimensional difference between groups since it is assumed that the treatment is unconfounded given this PS [Schafer and Kang, 2008].

2.2.4 Inverse Probability Weighting (IPW)

With the aforementioned PS, IPW is calculated by $\frac{1}{PS}$ for the vaccinated group and $\frac{1}{1-PS}$ for the unvaccinated. IPW creates a pseudo-population by equaling the effect of the confounders mimicking a random control trial [Shiba and Kawahara, 2021].

Two logistic regression models, with stabalized- and unstabilized weights were fitted. When using stabilized weights the numerator model also includes confounders, giving more stable estimates compared to the unstabilized weights [van der Wal and Geskus, 2011]. In both models bootstrapping was used to account for the inflated sample size of the pseudo population.

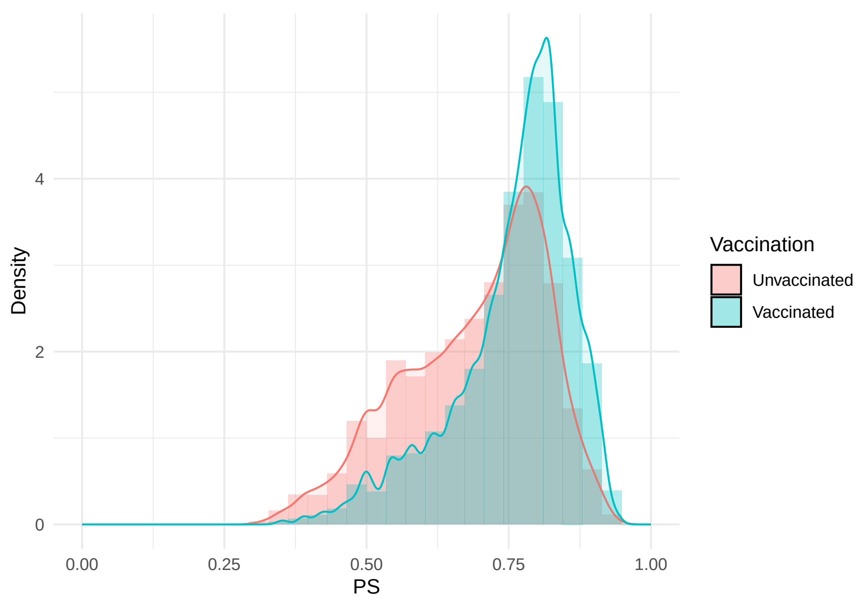

Figure 2: Distribution of the Propensity score for participants who received a vaccination compared with those who did not receive a vaccination

3 Results

Table 2 shows the crude association between vaccination status and hospitalization. The odds ratio (OR) was found to be 0.921 (95% CI: 0.699, 1.214), indicating a non-significant association (p = 0.562). The C-statistic, a measure of discrimination, was 0.508, suggesting poor predictive performance of the model.

Considering the second model in Table 2, which presents the PS as covariate, the OR suggests a significant negative association ($OR_{adjusted}$: 0.635. 95% CI 0.478 to 0.843; p <.001). The C-Statistic of 0.679 suggests that the model has moderate discriminatory power.

The IPW model with unstabilized weights yields significant negative association between the exposure and outcome variable ($OR_{adjusted}$: 0.617, 95%CI 0.523 to 0.728; p <.001). The model presents moderate discriminatory power, where C = 0.573.

The IPW model with stabilized weights. This model presents a negative association between vaccination and hospitalization ($OR_{adjusted}$: 0.617, 95% CI 0.509 to 0.748; p <.001). The C-statistic suggests moderate discriminatory power (C= 0.573).

Table 2: Association between influenza vaccination and hospitalization (n=40000)

Model Specification	OR (95% CI)	P-Value	C-Statistic
Unadjusted	0.921 (0.699 to 1.214)	0.562	0.508
PS as covariate	0.635 (0.478 to 0.843)	<.001	0.679
Unstabilized IPW	0.617 (0.523 to 0.728)	<.001	0.573
stabilized IPW	0.617 (0.509 to 0.748)	<.001	0.573

4 Discussion

The results indicate that the model with the PS as covariate gives the most clear results. Both IPW models have narrower confidence interval, however their C-statistic is very poor. This makes the PS as covariate model the most adequate. Accounted for confounders, this would mean that being vaccinated causes a lower odds of being hospitalized compared to people without a vaccination.

Other modelling methods could yield even more accurate results. Using machine learning techniques to calculate the PS could create an even better, overfitted, model [Schafer and Kang, 2008]. Using other techniques, like matching or a doubly robust estimation like weighted residual bias correction could also improve the results of the estimation. In this study it was chosen to only use the PS as a covariate and the IPW to compare the most common PS methods [Schafer and Kang, 2008].

Word Count: 800

References

Bahar Behrouzi, Deepak L Bhatt, Christopher P Cannon, Orly Vardeny, Douglas S Lee, Scott D Solomon, and Jacob A Udell. Association of influenza vaccination with cardiovascular risk: a meta-analysis. *JAMA Network Open*, 5(4):e228873–e228873, 2022.

Nederlands Huisartsen Genootschap. Risico hart- en vaatziekten verlagen. URL https://www.thuisarts.nl/risico-hart-en-vaatziekten-verlagen.

Martin Loef and Harald Walach. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. *Preventive medicine*, 55(3):163–170, 2012.

Joseph L Schafer and Joseph Kang. Average causal effects from nonrandomized studies: a practical guide and simulated example. *Psychological methods*, 13(4):279, 2008.

Koichiro Shiba and Takuya Kawahara. Using propensity scores for causal inference: pitfalls and tips. *Journal of epidemiology*, 31(8):457–463, 2021.

Yuxi Tian, Elande Baro, and Rongmei Zhang. Performance evaluation of regression splines for propensity score adjustment in post-market safety analysis with multiple treatments. *Journal of Biopharmaceutical Statistics*, 29(5):810–821, 2019.

Willem M. van der Wal and Ronald B. Geskus. ipw: An r package for inverse probability weighting. *Journal of Statistical Software*, 43(13):1–23, 2011. doi: 10.18637/jss.v043.i13. URL https://www.jstatsoft.org/index.php/jss/article/view/v043i13.

Daniel Westreich and Stephen R Cole. Invited commentary: positivity in practice. *American journal of epidemiology*, 171(6):674–677, 2010.