

# Assignment

Martijn Koster, 6234119      Jurrian van de Kraats, student nr  
Tim Poorthuis, student nr

2023-02-15

# 1 Introduction

In this paper the

In this paper we will examine whether there is a causal effect from elderly people (i.e., 65+) who have received an influenza vaccine and the likelihood of hospitalization. This will be examined by using the patient records dataset.

Write here a short motivation with a RQ

- well articulated RQ

**The aim of the study is to assess whether annual influenza vaccination reduces the risk of hospitalization among elderly (i.e., people aged  $\geq 65$  years).**

## 2 Methods

### 2.1 Examining 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’ was given as a proxy for the unobserved variable healthy lifestyle. An influenza vaccination is seasonal, hence does the time from last contact with the chiropractor form a causal path to vaccination status. Furthermore is the national distribution of an influenza vaccination determined by age, hence does age form a causal path towards vaccination status. Having obtained an influenza vaccination was associated with a lower risk for adverse cardiovascular events with hospitalization as a result [Behrouzi et al., 2022]. Having pulmonary disease and diabetes increases one’s chances to get severe cardiovascular disease [Genootschap]. The healthiness of one’s lifestyle forms a major cause in getting (severe) cardiovascular disease, pulmonary disease and diabetes. In return is once sex an important confounder in the healthiness of one’s lifestyle [Loef and Walach, 2012]. The previously mentioned diseases could all be a cause for hospitalization.

The causal paths of the DAG should be examined to assess whether unbiased causal inference is possible. The main two causal paths from vaccination status towards hospitalisation are a direct path and a path through the mediator cardiovascular disease. However, there is a flow of statistical information through open backdoor paths as well. This is due to observed and unobserved confounders like age, contact with chiropractor and healthy lifestyle. The eventual model would need to adjust for these confounders in order to perform an unbiased causal inference.

### 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, cardiovascular disease, pulmonary disease, diabetes and hospitalization were binomial. Table 1 presents the baseline characteristics of 40000 individuals who were included in the study. The mean age was 75.65 (sd = 6.97). The majority of the study were woman (61.91%). Approximately half (49.39%) of the respondents had Cardiovascular disease, whereas 12.34 percent had Pulmonary disease. Of the participants, 6.54% had Diabetes Mellitus.

A total of 29616 (74.04%) respondents were vaccinated. People who received a vaccination had on average more often pulmonary disease (14.33% vs 6.67%), cardiovascular disease (53.02% vs 39.05%) and diabetes (7.51% vs 3.8%).

Baseline characteristics stratified by the study outcome indicate that 254 of the respondents were hospitalized as opposed to 39746 who were not. Respondents who were hospitalized were older (79.68 vs 75.63), had

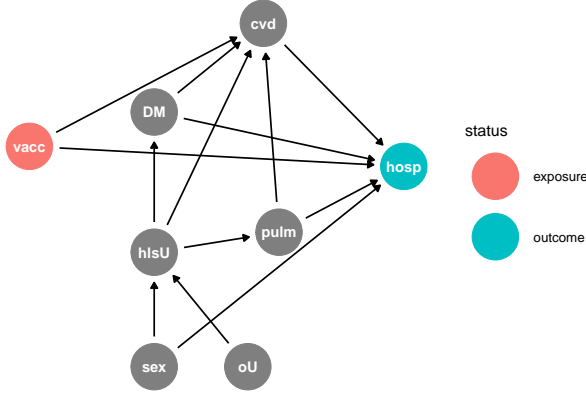


Figure 1: DAG model with Vaccine as exposure and hospitalization as outcome

more often contact with the GP (21.08 vs 14.71), were on average less often a female (48.43% vs 61.99%), had on average more often cardiovascular disease (72.05% vs 49.25%) and on average more often Diabetes (11.42% vs 6.51%).

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

Characteristics	Total	Influenza vaccination received		Hospitalized	
		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)
Cardiovascular disease, 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)

In order to control for confounding (observed and unobserved), a PS was estimated by fitting a logistic regression. A PS gives the probability being vaccinated for the respondents. The variables used to calculate the propensity score are derived from the aforementioned DAG (Figure 1), namely: age, sex, cardiovascular disease, pulmonary disease, diabetes mellitus, and GP contact in 12 months prior to start of study. For the variables age and contact a spline is used, because showed favorable performance compared to other propensity score methods [Tian et al., 2019].

In Figure 2, the PS for vaccinated and unvaccinated individuals appear to be well-balanced, indicating that the distribution of covariates between the two groups is similar. This plot also supports the positivity assumption, which means that both vaccinated and unvaccinated individuals are represented in all sub-populations defined by different combinations of covariates [Westreich and Cole, 2010].

### 2.2.3 PS as Covariate

The first adjusted model is to use the PS as a covariate. In an observational study the treated and untreated group have an equal distribution given that these are divided in groups of a constant propensity. Since it is assumed that the treatment is unconfounded given this propensity score, can the propensity score be used as a baseline variable to account for the dimensional difference between groups [Schafer and Kang, 2008]. This method also provides more statistical power and precision than matching, particularly when the sample size is large [Austin, 2011].

### 2.2.4 Inverse Probability Weighting (IPW)

With the aforementioned propensity scores, 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 replicating observations of individuals with higher weights, so the probability of exposure isn't affected by covariates. For example, if gender affects both exposure and outcome, IPW assigns greater weights to underrepresented groups, achieving a comparable gender ratio between the exposed and unexposed groups in the pseudo-population. This allows for comparison of the weighted outcome averages between the exposed and unexposed individuals in the pseudo-population, with no confounding from measured covariates [Shiba and Kawahara, 2021].

What's the difference compared to unstabilized weights? *The difference is in the numerator of the weights. Here, the results are equal, because the numerator is just a constant. However, the numerator model can include confounders too, in which case the stabilized weight yield more stable estimates.*

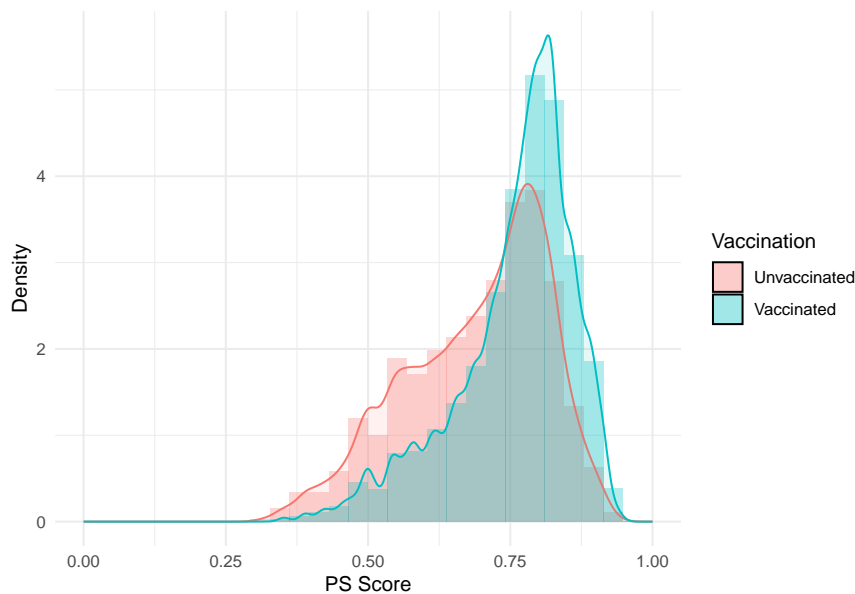


Figure 2: Distribution of the Propensity score for participants who received a vaccination compared with those who did not receive a vaccination

## 3 Results

As shown in Table 2, the crude association between vaccination status and hospitalization was examined using logistic regression, and the odds ratio 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 score as covariate, the Odds Ratio suggests a significant negative association (adjusted OR: 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 unstabilised weights yields significant negative association between the exposure and outcome variable (adjusted OR: 0.617, 95% CI 0.523 to 0.728;  $p < .001$ ). The model presents moderate discriminatory power, where  $C = 0.573$ .

The last model in Table 2 is an IPW model with stabilised weights. This model also presents a negative association between vaccination and hospitalisation (Adjusted OR: 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 score as covariate	0.635 (0.478 to 0.843)	<.001	0.679
Unstabilised IPW	0.617 (0.523 to 0.728)	<.001	0.573
Stabilised IPW	0.617 (0.509 to 0.748)	<.001	0.573

## 4 Conclusion / Discussion

- Conclusions supported by data
- Other issues (both positive and negative) Maybe write something about matching here?

## References

- Peter C Austin. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3):399–424, 2011.
- 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 Loeff 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, Elan 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.
- Daniel Westreich and Stephen R Cole. Invited commentary: positivity in practice. *American journal of epidemiology*, 171(6):674–677, 2010.