Global Epidemiology 6 (2023) 100120

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology

Commentary

Adjustment for collider bias in the hospitalized Covid-19 setting

Moslem Taheri Soodejani [a], Seyyed Mohammad Tabatabaei [b], Mohammad Hassan Lotfi [a], Maryam Nazemipour [c,*], Mohammad Ali Mansournia [c,*]

[a] Center for Healthcare Data Modeling, Departments of Biostatistics and Epidemiology, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
[b] Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
[c] Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Keywords
SARS-COV2
Vaccine
Effectiveness
Comorbidity

ABSTRACT

Background: Causal directed acyclic graphs (cDAGs) are frequently used to identify confounding and collider bias. We demonstrate how to use causal directed acyclic graphs to adjust for collider bias in the hospitalized Covid-19 setting.
Materials and methods: According to the cDAGs, three types of modeling have been performed. In model 1, only vaccination is entered as an independent variable. In model 2, in addition to vaccination, age is entered the model to adjust for collider bias due to the conditioning of hospitalization. In model 3, comorbidities are also included for adjustment of collider bias due to the conditioning of hospitalization in different biasing paths intercepting age and comorbidities.
Results: There was no evidence of the effect of vaccination on preventing death due to Covid-19 in model 1. In the second model, where age was included as a covariate, a protective role for vaccination became evident. In model 3, after including chronic diseases as other covariates, the protective effect was slightly strengthened.
Conclusion: Studying hospitalized patients is subject to collider-stratification bias. Like confounding, this type of selection bias can be adjusted for by inclusion of the risk factors of the outcome which also affect hospitalization in the regression model.

1. Background

Causal directed acyclic graphs (cDAGs) are frequently used to facilitate the study design, variable selection of statistical models for effect estimation, and creation of a graphical framework for classifying potential bias sources. cDAGs are graphical tools that posit the causal structure for the study population, in which the underlying causal relationships between exposure, outcome and other related variables are displayed using directed arrows [1–10].

cDAGs can be used to identify two sources of bias in the causal effect estimation: confounding and collider-stratification bias. Confounding arises from a common cause of the exposure and outcome; a confounder is any variable on this confounding path. In contrast, collider-stratification bias occurs due to conditioning on the collider which is a common effect of the exposure (or a cause of the exposure) and outcome (or a cause of the outcome) [11]. Both biases can be corrected by appropriate adjustment for a confounder and a non-collider variable on the collider-stratification biasing path, respectively.

During the Covid-19 pandemic, large databases have been used in epidemiological studies in Iran to evaluate the impact of various factors on the outcomes of Covid-19 [12–16]. The effect of vaccination/drugs on death are often studied in the hospitalized Covid-19 patients. Selecting samples from hospital information systems is one of the possible sources of collider-stratification bias, known as Berksonian bias [17]. In this manuscript, we demonstrate how to adjust for collider bias in the assessment of the effect of vaccination on death in the hospitalized Covid-19 setting using cDAGs.

2. Materials and methods

2.1. Data accessing

The design was cohort study restricted to Covid-19 patients hospitalized in Shahid Sadoughi Hospital, Yazd in 2020. The research

* Corresponding authors.
E-mail addresses: ma_nazemipour@yahoo.com (M. Nazemipour), mansournia_m@sina.tums.ac.ir (M.A. Mansournia).

https://doi.org/10.1016/j.gloepi.2023.100120
Received 15 June 2023; Received in revised form 14 August 2023; Accepted 22 August 2023
Available online 23 August 2023

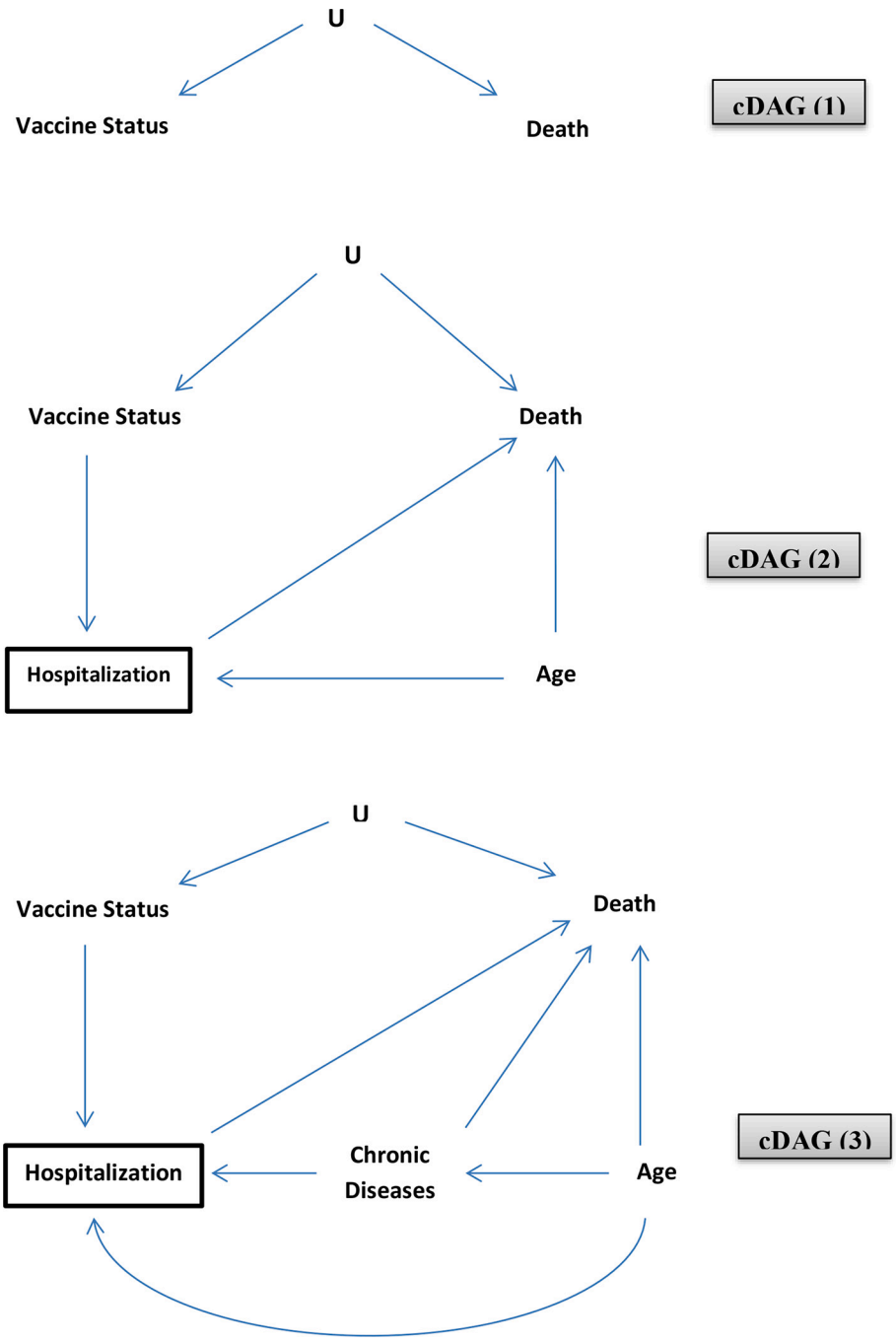

Fig. 1. Causal directed acyclic graphs outlining 3 scenarios under study (Chronic Diseases are Cancers, Diabetes Mellitus, Heart Diseases, Liver Diseases, and Kidney Diseases. U includes unmeasured factors that affect both vaccine status and death, e.g. personality type).

Table 1
Distribution of age and comorbidities in Covid-19 hospitalized by vaccine status.

Variables	Vaccine Status	
	Vaccine ($N = 4691$)	Non-vaccine ($N = 16,619$)
Age (years); mean (SD)	58.6 (18.6)	46.3 (21.2)
≥ 60	2440 (52%)	4495 (27%)
Cancers comorbidity	87 (1.9%)	103 (0.62%)
Liver diseases comorbidity	18 (0.38%)	11 (0.066%)
Heart diseases comorbidity	563 (12%)	429 (2.6%)
Diabetes mellitus comorbidity	1030 (22%)	1108 (6.7%)
Kidney diseases comorbidity	127 (2.7%)	91 (0.55%)

Data are No.(%) unless otherwise was specified.

question was the effect of vaccine status (two doses) on death in hospitalized Covid-19 patients. All patients referred to Covid-19 diagnostic centers were registered in medical care monitoring center (MCMC). Demographic data, underlying diseases such as heart diseases, kidney diseases, dementia, cancers, diabetes mellitus, and liver diseases registered based on ICD-10 codes were collected.

2.2. Causal directed acyclic graphs for the study structure

Fig. 1 shows cDAGs for three scenarios in this study. cDAG 1 represents relationship between Covid-19 vaccine status and death without considering any other measured variable. The variable U represents an unmeasured confounder such as personality type which affects both

Table 2
Three modeling approaches according to cDAGs in Fig. 1.

Variables	Model 1		Model 2		Model 3	
	OR(95% CI)	P-value	aOR*(95% CI)	P-value	aOR*(95% CI)	P-value
Vaccine Status	**1.0 (0.85–1.2)**	**0.91**	**0.65 (0.54–0.77)**	**<0.001**	**0.54 (0.45–0.66)**	**<0.001**
Age ≥ 60 years	–	–	6.1 (5.2–7.1)	<0.001	5.5 (4.6–6.4)	<0.001
Cancers comorbidity	–	–	–	–	3.1 (2.0–4.8)	<0.001
Liver diseases comorbidity	–	–	–	–	2.2 (0.71–7.0)	0.17
Heart diseases comorbidity	–	–	–	–	1.3 (1.0–1.7)	0.029
Diabetes mellitus comorbidity	–	–	–	–	1.4 (1.2–1.7)	0.001
Kidney diseases comorbidity	–	–	–	–	3.2 (2.2–4.7)	<0.001

* Adjusted odds ratio.

vaccine status and death. In cDAG 2, we added an arrow from age to death, and hospitalization was added as a collider, i.e., a common effect of vaccine status and age. The square around the variable hospitalization indicates that the analysis was restricted to hospitalized patients. In this case, a biasing path is opened from vaccine status to death due to conditioning on the variable hospitalization, and age should be adjusted for in the analysis. In cDAG 3, some chronic diseases as comorbidities were added to the cDAG. In this cDAG, conditioning on the variable hospitalization produces biasing paths from vaccine status to death, and so both variables age and chronic diseases should be adjusted for in the analysis. This is true even if age or chronic disease are not causal confounders, reflected in no arrow from age or chronic disease to vaccine status in cDAGs 2–3 i.e., the analysis was performed when the vaccine was available to all persons in the population.

2.3. Modeling and data analysis

According to the cDAGs in Fig. 1, three types of logistic regression model have been performed. In model 1, vaccine status was entered as an independent variable. In model 2, vaccine status and age were entered. In model 3, vaccine status, age, and comorbidities were included.

Model 1: $\text{logit}(\pi) = \beta_0 + \beta 1 \times \text{Vaccine status}$.

Model 2: $\text{logit}(\pi) = \beta_0 + \beta 1 \times \text{Vaccine status} + \beta_2 \text{Age}$.

Model 3: $\text{logit}(\pi) = \beta_0 + \beta_1 \times \text{Vaccine status} + \beta_2 \text{Age} + \beta_3 \times \text{Cancers comorbidity} + \beta_4 \times \text{Liver diseases comorbidity} + \beta_5 \times \text{Heart diseases comorbidity} + \beta_6 \times \text{Diabetes mellitus comorbidity} + \beta_7 \times \text{Kidney diseases comorbidity}$.

where π is the probability of death during hospitalization.

3. Results

21,310 confirmed case of Covid-19 using RT-PCR with a mean age of 49.0 ± 21.3 years were admitted from May 1 to November 1, 2020, of which, 4691 were fully vaccinated. Investigating their mean age showed that it was higher in those who died than in those who survived. Table 1 presents distribution of age and comorbidities in Covid-19 hospitalized by vaccine status.

Based on model 1, there was no evidence of the effect of vaccination on preventing death due to Covid-19: the results were compatible with both protective and risk effects (OR: 1.0, 95% CI: 0.85–1.2) [18,19]. In the second model, where age was included as a covariate, a protective role for vaccination became evident (aOR: 0.65, 95% CI: 0.54–0.77). At last, after including chronic diseases as other covariates, the protective effect was slightly strengthened (aOR: 0.54, 95% CI: 0.45–0.66). Table 2 provides more details of these three models. The results suggest that collider bias due to hospitalization masked the protective effect of vaccine status on death in Covid-19 patients.

4. Discussion

There was no evidence of relationship between vaccination and death due to Covid-19 in the hospitalized patients (Model 1). However, this crude analysis is subject to selection bias due to restriction to hospitalized patients. In fact, age positively and vaccine status negatively affect hospitalization i.e., old and/or unvaccinated subjects are likely hospitalized (Fig. 1, cDAG 2), and so a positive association between age and vaccine status among hospitalized patients is expected (Table 1) [11].

The latter together with the positive effect of age on death produce a spurious positive association between vaccine status and death which seems to cancel the protective (negative) effect of vaccine status on death which was confirmed in an age-adjusted analysis (Model 2). This is an example of collider-stratification bias due to conditioning on the collider hospitalization which can be corrected using adjustment for the measured variable age which lies on the collider biasing path.

A slightly strengthened protective effect estimate was obtained in Model 3 which further adjusted for comorbidities. Like age, comorbidities positively affects hospitalization and so using the arguments similar to those mentioned above, the bias due to no adjustment for comorbidities should be towards the null value. However, this bias was not substantial conditional on age for two reasons. First, the effects of all comorbidities on hospitalization may not be as strong as the effect of age. Second, age and comorbidities are positively associated and age can be considered as a proxy for comorbidities, and so adjustment for comorbidities when age has been already adjusted for is not expected to substantially change the effect estimate.

This study underscores the importance of collider-stratification bias when hospitalized patients are analyzed. Fortunately, like confounding, this type of selection bias can be adjusted for by inclusion of the risk factors of the outcome which also affect hospitalization in the regression model. It is important to note that unmeasured confounding may be still present (the variable U in Fig. 1).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Etminan M, Collins GS, Mansournia MA. Using causal diagrams to improve the design and interpretation of medical research. Chest 2020;158(1) (S21-S8).
[2] Kyriacou DN, Greenland P, Mansournia MA. Using Causal Diagrams for Biomedical Research. Annals of emergency medicine 2023;81(5):606–13.
[3] Etminan M, Brophy JM, Collins G, Nazemipour M, Mansournia MA. To adjust or not to adjust: the role of different covariates in cardiovascular observational studies. Am Heart J 2021;237:62–7.
[4] Etminan M, Nazemipour M, Mansournia MA. Potential biases in studies of acid-suppressing drugs and COVID-19 infection. Gastroenterology. 2021;160(5): 1443–6.
[5] Mansournia MA, Collins GS, Nielsen RO, Nazemipour M, Jewell NP, Altman DG, et al. A CHecklist for statistical Assessment of Medical Papers (the CHAMP statement): explanation and elaboration. Br J Sports Med 2021;55(18):1009–17.
[6] Mansournia MA, Collins GS, Nielsen RO, Nazemipour M, Jewell NP, Altman DG, et al. CHecklist for statistical Assessment of Medical Papers: the CHAMP statement. Br J Sports Med 2021;55(18):1002–3.

[7] Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. Int J Epidemiol 2013;42(3):860–9.
[8] Mansournia MA, Higgins JP, Sterne JA, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. Epidemiology. 2017;28(1):54.
[9] Mansournia MA, Nazemipour M, Etminan M. Causal diagrams for immortal time bias. Int J Epidemiol 2021;50(5):1405–9.
[10] Mansournia MA, Nazemipour M, Etminan M. Time-fixed vs time-varying causal diagrams for immortal time bias. Int J Epidemiol 2022;51(3):1030–1.
[11] Mansournia MA, Nazemipour M, Etminan M. Interaction contrasts and collider bias. Am J Epidemiol 2022;191(10):1813–9.
[12] Shakiba M, Nazemipour M, Salari A, Mehrabian F, Nazari SSH, Rezvani SM, et al. Seroprevalence of SARS-CoV-2 in guilan province, Iran, April 2020. Emerg Infect Dis 2021;27(2):636.
[13] Shakiba M, Nazemipour M, Heidarzadeh A, Mansournia M. Prevalence of asymptomatic COVID-19 infection using a seroepidemiological survey. Epidemiology Infection. 2020:148.
[14] Rostami A, Sepidarkish M, Fazlzadeh A, Mokdad AH, Sattarnezhad A, Esfandyari S, et al. Update on SARS-CoV-2 seroprevalence: regional and worldwide. Clinical Microbiology Infection 2021;27(12):1762–71.
[15] Doosti-Irani A, Mostafavi E, Nazemipour M, Mansournia MA, Haghdoost A-A. Challenges for management of the COVID-19 epidemic in Iran. Global epidemiology. 2020;2:100035.
[16] Arman A, Tajik M, Nazemipour M, Ahmadinejad Z, Shahrestanaki SK, Hazrati E, et al. Risk factors of developing critical conditions in Iranian patients with COVID-19. Global epidemiology 2021;3:100046.
[17] Berkson J. Limitations of the application of fourfold table analysis to hospital data. Biometrics. 1946;2(3):47–53.
[18] Mansournia MA, Nazemipour M, Etminan M. P-value, compatibility, and S-value. Global Epidemiology 2022;4:100085.
[19] Greenland S, Mansournia MA, Joffe M. To curb research misreporting, replace significance and confidence by compatibility: a Preventive Medicine golden jubilee article. Prev Med 2022;107127.