

「2023년도 코로나19백신안전성및평가연구센터운영」  
전체사망(코로나 사망)과 백신접종률과의 관계  
분석 결과보고서

2023.11.30.



**코로나19백신안전성연구센터**

COVID-19 Vaccine Safety Research Center



**대한민국 의학한림원**

National Academy of Medicine of Korea

# 전체사망(코로나 사망)과 백신접종률과의 관계

## ■ abstract(요약)

- ▶ 코로나19백신 접종 후 사망과 백신 접종과의 상관성을 평가하기 위한 연구를 수행함.
- ▶ 총 2가지 2SLS를 활용한 연구(1. 전체사망, 2. 코로나 사망)에서 코로나19 1 ~ 4차 백신접종 후 전체사망 및 코로나 사망과의 명백한 증거를 발견하지 못 함.
- ▶ 추가로, 민감도 분석으로 Granger Causality test를 시행하였으나 코로나19 1 ~ 4차 백신접종 후 전체사망 및 코로나 사망과의 시계열적 증거를 발견하지 못 함.

*Keywords: 코로나19, 코로나19백신*

## 1. 분석방법론

### 1) Two-Step Least Square (2SLS)

#### (1) 연구방법론

##### ① 회귀분석의 기본가정

- 통제변수가 있는 경우 독립변수(X)와 오차항(error term)의 상관관계가 없어야 함
- 하지만 완벽하게 통제된 실험연구가 아닌 2차 데이터를 수집 및 분석하는 경우에는 위의 조건을 충족시키기 어려움
- 일반적으로 실험군에서 treatment가 없었을 때 의미하는 counter actual과 대조군의 차이를 선택적 편향(selection bias)라고 함
- 관찰되지 않은 요인에 의한 selection bias는 모두 회귀식의 오차항에 담김
- 즉, 오차항은 종속변수(Y)에는 영향을 미치지만 분석 중 관찰되지 못한 요인들은 모두 오차항에 포함됨
- 따라서 selection bias가 오차항에 포함되어 있기 때문에 오차항은 treatment와 상관관계를 가짐
- 독립변수와 오차항이 상관관계가 있다는 것은 selection bias가 있음을 의미하고 이러한 상관관계가 있을 경우 회귀분석의 오류를 범함

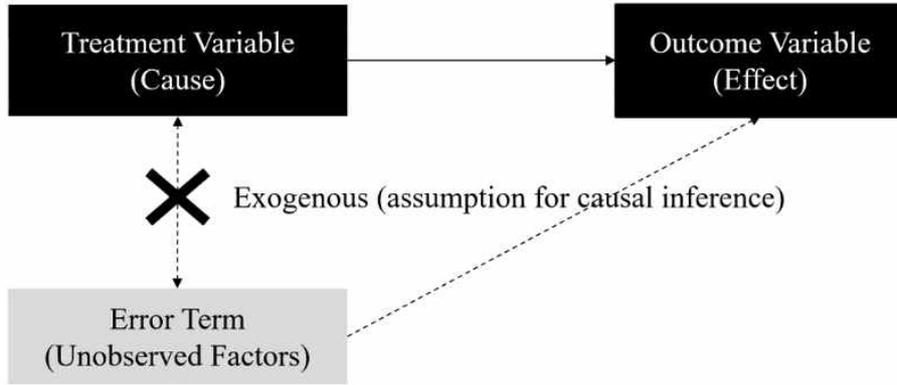


그림 1. Endogeneity in Regression

자료출처: Korea Summer Workshop on Causal Inference 2022

- 그림 1처럼 treatment variable과 오차항이 exogenous한 관계일 경우 인과관계를 설명할 수 있지만 실제 사이에 endogenous한 요소가 섞여 있을 경우 명확한 인과추론이 어려움
- 따라서 selection bias와 같은 endogeneity 문제를 해결하기 위해서는 도구변수(Instrument variable, IV)를 활용한 Two-Step Least Square (2SLS)방법론을 제시함

② 도구변수를 활용한 2SLS

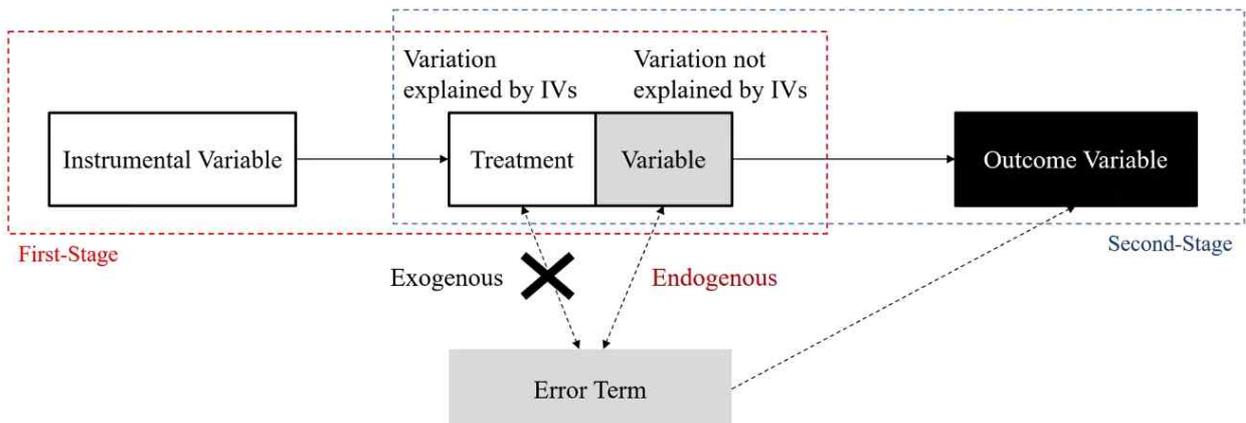


그림 2. Instrument Variable

자료출처: Korea Summer Workshop on Causal Inference 2022

- 도구변수는 그림 2처럼 exogeneous와 endogeneous한 요소를 분리하는 도구적인 역할을 수행하며 도구변수를 활용하여 exogeneous한 부분만 예측가능하다면 treatment variable이 종속변수에 미치는 인과관계를 해석할 수 있음
- 그리고 독립변수에서 도구변수로 설명되지 않는 부분이 selection bias에 해당됨

### ③ 도구변수의 조건

- 일반 회귀식

$$y = \beta_0 + \beta_1 x + u \quad \text{cov}(x, u) \neq 0$$

→ x와 u 사이에 상관관계가 있으면 x는 endogenous variable임

- First step

$$x = \beta_0 + \beta_1 z + \nu$$

$$\text{where } \beta_0 \neq 0, \text{ cov}(z, u) = 0, \text{ and } \text{cov}(z, \nu) = 0$$

→ x는 독립변수, z는 도구변수,  $\nu$ 는 오차항 임

→ 도구변수는 일반 회귀식의 y에 영향을 미치지만 오차항인 u와는 상관관계가 없어야 함

- Second step

$$\hat{x} = \beta_0 + \beta_1 z$$

$$y = \gamma_0 + \gamma_1 \hat{x} + u$$

$$\text{where } \text{cov}(\hat{x}, u) = \text{cov}(\beta_0 + \beta_1 z, u) = 0$$

→  $\hat{x}$ 은 도구변수로 예측한 exogenous 부분임

→  $\hat{x}$ 변수는 내생성 문제가 없으므로 오차항과의 상관관계는 0임

## 2. 연구설계

### 1) 변수 정의

- 종속변수: 일별 전체 사망자 수(코로나 사망 제외)
- 독립변수: 1차, 2차, 3차, 4차 백신접종 완료자 수
- 도구변수: 사회적 거리두기 점수(Social Distancing Score, SDS)
  - 유동인구가 많은 우리나라 주요 도시 11개소 지하철역 선정
    - 서울: 명동, 신촌, 홍대입구, 강남역
    - 부산: 남포, 광안, 연산역
    - 대구: 서문시장, 중앙로역
    - 대전: 중앙로역
    - 광주: 광주송정역

### S-SDS

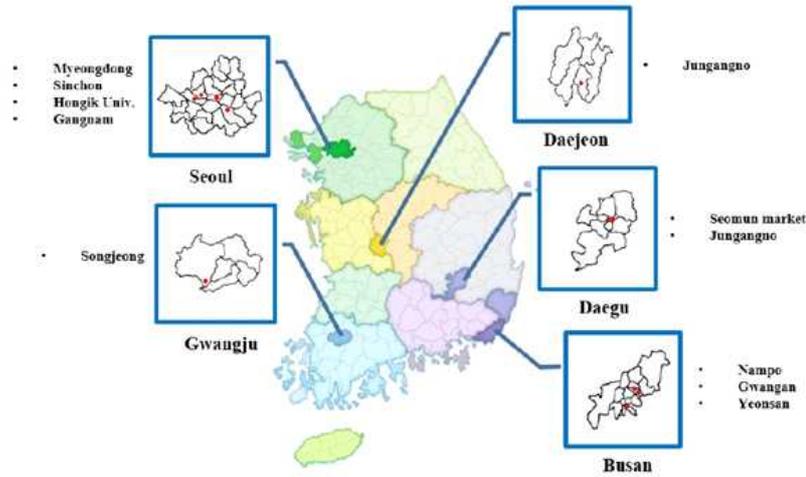


그림 3. Subway Stations for Social Distancing Scores

- SDS 점수화 산식<sup>1)</sup>

→ 주별 지하철역 유동인구 수 / 지난 4년간(2020년 이전) 주별 평균 지하철역 유동인구 수

$$\text{Subway use - based SDS} = \frac{\text{The number of passengers at the subway stations per week}}{\text{The average number of weekly passengers in the last 4 years}}$$

### 2) 연구모형

$$All\ Death = \alpha + \beta_1 Confirmed\ COVID-19 + \beta_2 Vaccine + \beta_3 POP + \beta_4 SDS + \varepsilon$$

→ All Death: 전체 사망자 수, Confirmed COVID-19: 코로나 확진자 수, Vaccine: n차 백신 접종 완료자 수, POP: 전체 인구 수, SDS: 사회적 거리두기 점수,  $\varepsilon$ : 오차항

→ 추정된  $\hat{\beta}_4$ 을 위의 회귀식에 적용

- 최종연구모형

$$All\ Death = \alpha + \beta_1 Confirmed\ COVID-19 + \beta_2 Vaccine + \beta_3 POP + \hat{\beta}_4 SDS + \varepsilon$$

- 추가적으로 분석 시 최종연구모형에 계절성까지 고려함

1) Seong H, Hong J-W, Hyun H-J, Yoon J-G, Noh J-Y, Cheong H-J, Kim W-J, Jung J-H, Song J-Y. Correlation between the Level of Social Distancing and Activity of Influenza Epidemic or COVID-19 Pandemic: A Subway Use-Based Assessment. *Journal of Clinical Medicine*. 2021; 10(15):3369. <https://doi.org/10.3390/jcm10153369>

### 3) 자료원 구축

표 1. 자료원 정보

1	질병관리청 코로나19 접종 등록정보(2021.02.26.~2022.12.31.)와 국민건강보험 청구자료 연계 자료원(2019.01.01.~2022.12.31.)
2	선정기준 - 코로나19 접종 등록정보 기준 2021년 2월 27일부터 2022년 12월 31일까지 접종 이력 존재자
3	제외기준 - 백신 접종 이력이 4차까지 없는 경우

### 3. 연구결과

#### 1) 1 ~ 4차 누적 백신접종자 수와 전체 사망자 수(코로나 사망 제외)

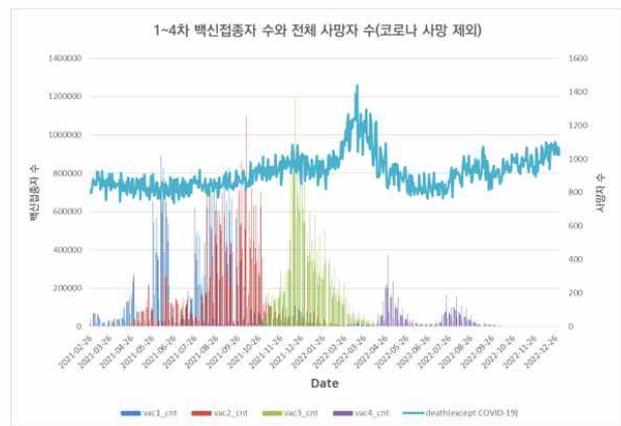
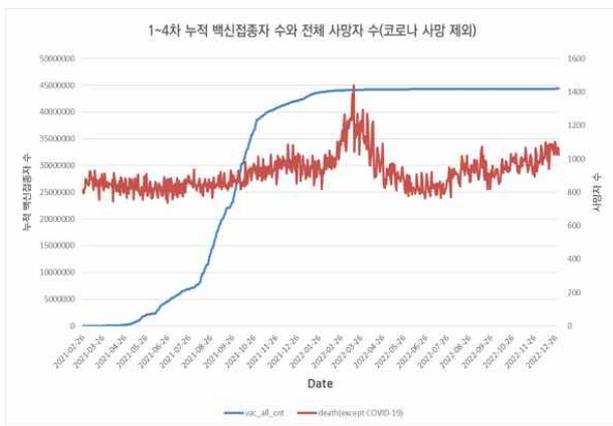


그림 4. 1 ~ 4차 (누적)백신접종자 수와 전체 사망자 수(코로나 사망 제외)

○ **(요약)** 코로나19백신 1 ~ 4회 누적 접종자 수는 2022년 2월부터 약 4,500만명이며, 동일 기간 사망자 수는 평균 1,000명을 보였다. 그리고 2022년 3월말 경 전체 사망자 수가 일시적으로 1,400명이 관찰되었다.



그림 5. 전체 사망자 수(코로나 사망 제외)와 사회적 거리두기 점수

○ **(요약)** 2020년 2월 29일부터 2022년 4월 18일까지 사회적 거리두기 정책이 시행되었으며, 사회적 거리두기 정책 시행동안 지하철역 승하차인원으로 SDS를 도출하였다.

(1) 2SLS을 활용한 1차 백신접종과 전체 사망자 수와의 연관성

○ **(요약)** 1차 백신접종과 전체 사망자 수를 대상으로 OLS 회귀분석 결과, 두 변수사이의 상관성이 매우 높았으나 SDS를 도구변수로 포함하여 분석한 2SLS 분석결과에서는 통계적 유의성이 관찰되지 않았다.

표 2. 1차 백신접종과 전체 사망자 수와의 연관성

구분		estimated value	P-value
코로나19백신 1차 접종	OLS	-1.480	<.0001
	2SLS	-1.574	0.8930

(2) 2SLS을 활용한 2차 백신접종과 전체 사망자 수와의 연관성

○ **(요약)** 2차 백신접종과 전체 사망자 수를 대상으로 OLS 회귀분석 결과, 두 변수사이의 상관성이 매우 높았으나 SDS를 도구변수로 포함하여 분석한 2SLS 분석결과에서는 통계적 유의성이 관찰되지 않았다.

표 3. 2차 백신접종과 전체 사망자 수와의 연관성

구분		estimated value	P-value
코로나19백신 2차 접종	OLS	-9.979	<.0001
	2SLS	-1.811	0.8930

### (3) 2SLS을 활용한 3차 백신접종과 전체 사망자 수와의 연관성

- **(요약)** 3차 백신접종과 전체 사망자 수를 대상으로 OLS 회귀분석 결과, 두 변수사이의 상관성이 매우 높았으나 SDS를 도구변수로 포함하여 분석한 2SLS 분석결과에서는 통계적 유의성이 관찰되지 않았다.

표 4. 3차 백신접종과 전체 사망자 수와의 연관성

구분		estimated value	P-value
코로나19백신 3차 접종	OLS	2.176	<.0001
	2SLS	1.225	0.6520

### (4) 2SLS을 활용한 4차 백신접종과 전체 사망자 수와의 연관성

- **(요약)** 4차 백신접종과 전체 사망자 수를 대상으로 OLS 회귀분석 결과, 두 변수사이의 상관성이 없었으며 SDS를 도구변수로 포함하여 분석한 2SLS 분석결과에서도 통계적 유의성이 관찰되지 않았다.

표 5. 4차 백신접종과 전체 사망자 수와의 연관성

구분		estimated value	P-value
코로나19백신 4차 접종	OLS	-4.477	0.7440
	2SLS	-3.188	0.6600

### (5) 2SLS을 활용한 1 ~ 4차 누적백신접종과 전체 사망자 수와의 연관성

- **(요약)** 1 ~ 4차 누적백신접종과 전체 사망자 수를 대상으로 OLS 회귀분석 결과, 두 변수사이의 상관성이 매우 높았으나 SDS를 도구변수로 포함하여 분석한 2SLS 분석결과에서는 통계적 유의성이 관찰되지 않았다.

표 6. 1 ~ 4차 누적백신접종과 전체 사망자 수와의 연관성

구분		estimated value	P-value
코로나19백신 1~4차 누적 접종	OLS	4.003	<.0001
	2SLS	1.323	0.8920

2) 1 ~ 4차 누적 백신접종자 수와 코로나 사망자 수

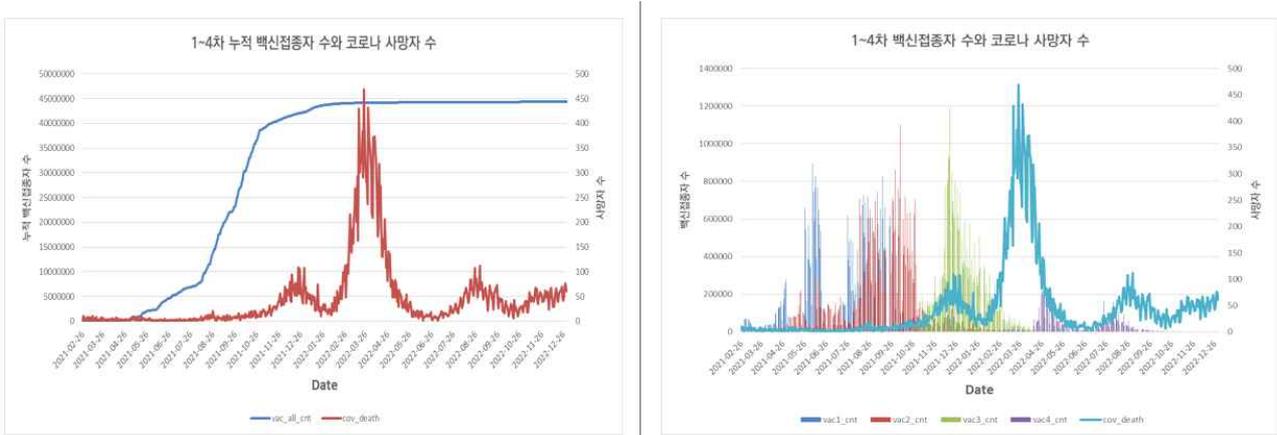


그림 6. 1~4차 누적 백신접종자 수와 코로나 사망자 수

- (요약) 코로나19백신 1 ~ 4회 누적 접종자 수는 2022년 2월부터 약 4,500만명이며, 동일 기간 코로나 사망자 수는 평균 43명을 보였다. 그리고 2022년 3월말 경 코로나 사망자 수가 최대 469명이 관찰되었다.

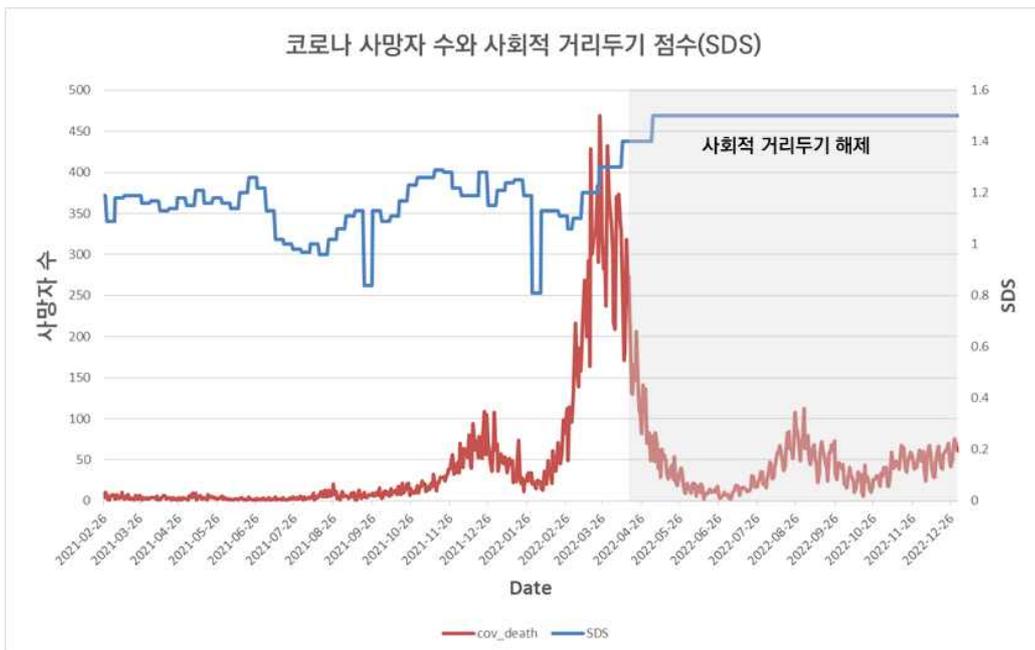


그림 7. 코로나 사망자 수와 사회적 거리두기 점수

- (요약) 2020년 2월 29일부터 2022년 4월 18일까지 사회적 거리두기 정책이 시행되었으며,

사회적 거리두기 정책 시행동안 지하철역 승하차인원으로 SDS를 도출하였다.

### (1) 2SLS을 활용한 1차 백신접종과 코로나 사망자 수와의 연관성

- **(요약)** 1차 백신접종과 코로나 사망자 수를 대상으로 OLS 회귀분석 결과, 두 변수사이의 상관성이 매우 높았으나 SDS를 도구변수로 포함하여 분석한 2SLS 분석결과에서는 통계적 유의성이 관찰되지 않았다.

표 7. 1차 백신접종과 코로나 사망자 수와의 연관성

구분		estimated value	P-value
코로나19백신 1차 접종	OLS	-6.287	<.0001
	2SLS	-1.009	<b>0.0864</b>

### (2) 2SLS을 활용한 2차 백신접종과 코로나 사망자 수와의 연관성

- **(요약)** 2차 백신접종과 코로나 사망자 수를 대상으로 OLS 회귀분석 결과, 두 변수사이의 상관성이 매우 높았으나 SDS를 도구변수로 포함하여 분석한 2SLS 분석결과에서는 통계적 유의성이 관찰되지 않았다.

표 8. 2차 백신접종과 코로나 사망자 수와의 연관성

구분		estimated value	P-value
코로나19백신 2차 접종	OLS	-5.795	<.0001
	2SLS	-1.161	<b>0.0887</b>

### (3) 2SLS을 활용한 3차 백신접종과 코로나 사망자 수와의 연관성

- **(요약)** 3차 백신접종과 코로나 사망자 수를 대상으로 OLS 회귀분석 결과, 두 변수사이의 상관성이 매우 높았으나 SDS를 도구변수로 포함하여 분석한 2SLS 분석결과에서는 통계적 유의성이 관찰되지 않았다.

표 9. 3차 백신접종과 코로나 사망자 수와의 연관성

구분		estimated value	P-value
코로나19백신 3차 접종	OLS	3.417	0.0178
	2SLS	2.282	<b>0.1120</b>

### (4) 2SLS을 활용한 4차 백신접종과 코로나 사망자 수와의 연관성

- **(요약)** 4차 백신접종과 코로나 사망자 수를 대상으로 OLS 회귀분석 결과, 두 변수사이의 상관성이 없었으며 SDS를 도구변수로 포함하여 분석한 2SLS 분석결과에서도 통계적 유의성이 관찰되지 않았다.

표 10. 4차 백신접종과 코로나 사망자 수와의 연관성

구분	estimated value	P-value
코로나19백신 4차 접종	OLS	1.139
	2SLS	5.940
		<b>0.1030</b>

(5) 2SLS을 활용한 1 ~ 4차 누적 백신접종과 코로나 사망자 수와의 연관성

- **(요약)** 1 ~ 4차 누적 백신접종과 코로나 사망자 수를 대상으로 OLS 회귀분석 결과, 두 변수사이의 상관성이 매우 높았으며 SDS를 도구변수로 포함하여 분석한 2SLS 분석결과에서는 통계적 유의성이 관찰되지 않았다.

표 11. 1 ~ 4차 누적 백신접종과 코로나 사망자 수와의 연관성

구분	estimated value	P-value
코로나19백신 1~4차 누적접종	OLS	1.209
	2SLS	-8.482
		<b>0.0844</b>

3) 민감도 분석

(1) Granger Causality를 활용한 1차 백신접종과 전체 사망자 수와의 연관성

- **(요약)** 1차 백신접종과 전체 사망자 수와의 관련성을 Ganger Causality test를 통해 최대 10일까지 지연효과를 고려하여 확인해 본 결과, 두 변수 간 관련성이 나타나지 않음

	전체 사망 -> 1차 백신접종 F-통계량	1차 백신접종 -> 전체 사망 F-통계량
Lag 1 day	0.7868	0.3430
Lag 2 days	0.4838	0.7241
Lag 3 days	0.9610	0.8159
Lag 4 days	0.5346	0.8054
Lag 5 days	0.9335	0.2993
Lag 6 days	0.9146	0.5619
Lag 7 days	0.1320	0.6172
Lag 8 days	0.0677	0.8792
Lag 9 days	0.0590	0.5143
Lag 10 days	0.6166	0.8614

(2) Granger Causality를 활용한 2차 백신접종과 전체 사망자 수와의 연관성

- **(요약)** 2차 백신접종과 전체 사망자 수와의 관련성을 Ganger Causality test를 통해 최대

10일까지 지연효과를 고려하여 확인해 본 결과, 두 변수 간 관련성이 나타나지 않음

	전체 사망 -> 2차 백신접종 F-통계량	2차 백신접종 -> 전체 사망 F-통계량
Lag 1 day	0.3693	0.9696
Lag 2 days	0.4703	0.3509
Lag 3 days	0.7858	0.6906
Lag 4 days	0.2713	0.3296
Lag 5 days	0.2533	0.1787
Lag 6 days	0.1023	0.3635
Lag 7 days	0.1213	0.3721
Lag 8 days	0.1998	0.4393
Lag 9 days	0.4946	0.7186
Lag 10 days	0.8719	0.7841

### (3) Granger Causality를 활용한 3차 백신접종과 전체 사망자 수와의 연관성

- **(요약)** 3차 백신접종과 전체 사망자 수와의 관련성을 Ganger Causality test를 통해 최대 10일까지 지연효과를 고려하여 확인해 본 결과, 두 변수 간 관련성이 나타나지 않음

	전체 사망 -> 3차 백신접종 F-통계량	3차 백신접종 -> 전체 사망 F-통계량
Lag 1 day	0.4323	0.7534
Lag 2 days	0.6084	0.5456
Lag 3 days	0.2751	0.3636
Lag 4 days	0.4668	0.4982
Lag 5 days	0.2055	0.7327
Lag 6 days	0.6458	0.0670
Lag 7 days	0.0591	0.1375
Lag 8 days	0.4054	0.8278
Lag 9 days	0.8220	0.4636
Lag 10 days	0.9666	0.4818

### (4) Granger Causality를 활용한 4차 백신접종과 전체 사망자 수와의 연관성

- **(요약)** 4차 백신접종과 전체 사망자 수와의 관련성을 Ganger Causality test를 통해 최대 10일까지 지연효과를 고려하여 확인해 본 결과, 두 변수 간 관련성이 나타나지 않음

	전체 사망 -> 4차 백신접종 F-통계량	4차 백신접종 -> 전체 사망 F-통계량
Lag 1 day	0.4128	0.8610
Lag 2 days	0.0548	0.1145
Lag 3 days	0.8536	0.0537
Lag 4 days	0.5717	0.2506
Lag 5 days	0.2218	0.0538
Lag 6 days	0.9037	0.1809
Lag 7 days	0.0505	0.3066
Lag 8 days	0.1948	0.4127
Lag 9 days	0.6817	0.2484
Lag 10 days	0.3389	0.0831

(5) Granger Causality를 활용한 1차 백신접종과 코로나 사망자 수와의 연관성

- (요약) 1차 백신접종과 코로나 사망자 수와의 관련성을 Ganger Causality test를 통해 최대 10일까지 지연효과를 고려하여 확인해 본 결과, 두 변수 간 관련성이 나타나지 않음

	코로나 사망 -> 1차 백신접종 F-통계량	1차 백신접종 -> 코로나 사망 F-통계량
Lag 1 day	0.8161	0.9016
Lag 2 days	0.8799	0.8686
Lag 3 days	0.7955	0.9855
Lag 4 days	0.6418	0.5397
Lag 5 days	0.9134	0.6255
Lag 6 days	0.8187	0.9209
Lag 7 days	0.7220	0.9478
Lag 8 days	0.7547	0.7580
Lag 9 days	0.7177	0.9901
Lag 10 days	0.8109	0.7497

(6) Granger Causality를 활용한 2차 백신접종과 코로나 사망자 수와의 연관성

- (요약) 2차 백신접종과 코로나 사망자 수와의 관련성을 Ganger Causality test를 통해 최대 10일까지 지연효과를 고려하여 확인해 본 결과, 두 변수 간 관련성이 나타나지 않음

	코로나 사망 -> 2차 백신접종 F-통계량	2차 백신접종 -> 코로나 사망 F-통계량
Lag 1 day	0.9857	0.9319
Lag 2 days	0.8423	0.9218
Lag 3 days	0.7675	0.6724
Lag 4 days	0.9968	0.4684
Lag 5 days	0.9876	0.8174
Lag 6 days	0.9037	0.8686
Lag 7 days	0.6275	0.9313
Lag 8 days	0.5838	0.7415
Lag 9 days	0.8715	0.9169
Lag 10 days	0.5340	0.7405

(7) Granger Causality를 활용한 3차 백신접종과 코로나 사망자 수와의 연관성

○ (요약) 3차 백신접종과 코로나 사망자 수와의 관련성을 Ganger Causality test를 통해 최대 10일까지 지연효과를 고려하여 확인해 본 결과, 두 변수 간 관련성이 나타나지 않음

	코로나 사망 -> 3차 백신접종 F-통계량	3차 백신접종 -> 코로나 사망 F-통계량
Lag 1 day	0.2903	0.9946
Lag 2 days	0.8087	0.6311
Lag 3 days	0.6015	0.3601
Lag 4 days	0.9715	0.7670
Lag 5 days	0.9926	0.4667
Lag 6 days	0.9129	0.0692
Lag 7 days	0.8426	0.3469
Lag 8 days	0.8680	0.5711
Lag 9 days	0.9135	0.9851
Lag 10 days	0.9231	0.3291

(8) Granger Causality를 활용한 4차 백신접종과 코로나 사망자 수와의 연관성

○ (요약) 4차 백신접종과 코로나 사망자 수와의 관련성을 Ganger Causality test를 통해 최대 10일까지 지연효과를 고려하여 확인해 본 결과, 두 변수 간 관련성이 나타나지 않음

	코로나 사망 -> 4차 백신접종 F-통계량	4차 백신접종 -> 코로나 사망 F-통계량
Lag 1 day	0.2764	0.0583
Lag 2 days	0.3005	0.3156
Lag 3 days	0.6115	0.5523
Lag 4 days	0.6719	0.2693
Lag 5 days	0.4702	0.4917
Lag 6 days	0.1039	0.6487
Lag 7 days	0.3094	0.9743
Lag 8 days	0.6156	0.0764
Lag 9 days	0.2693	0.3324
Lag 10 days	0.8102	0.1267

#### 4. 결론

- 대한민국 의학한림원 코로나19백신 안전성 위원회 역학 1분과는 코로나19백신 접종 후 사망과 백신 접종과의 상관성을 평가하기 위한 연구를 수행하였다.
- 총 2가지 2SLS를 활용한 연구(1. 전체사망, 2. 코로나 사망)에서 코로나19 1 ~ 4차 백신접종 후 전체사망 및 코로나 사망과의 명백한 증거를 발견하지 못했다. 본 연구결과는 도구변수를 고려하여 분석 시 발생할 수 있는 선택적 편향을 통제한 후 시행되었다.
- 또한 민감도 분석을 위해 Granger Causality test를 시행한 결과 코로나19 1 ~ 4차 백신접종 후 전체사망 및 코로나 사망과의 시계열적 증거를 발견하지 못했다. 본 연구결과는 백신접종 후 최대 10일까지의 지연효과를 고려하여 분석하였다.
- 그러나 본 연구는 다음과 같은 한계를 가진다.
- 도구변수를 고려할 때 사회적 거리두기 점수(SDS) 이외에 더 적합한 변수를 찾아 적용하는 연구가 반드시 요구된다.
- 시계열 데이터를 기반으로 한 Granger Causality test는 시차(lag)에 매우 민감하게 반응함으로 적합한 시차를 선택하는 것이 매우 중요하다.