



KULSOM

Track A 논문팀 최종 결과 발표

인과 추론 기법과 논문 적용

2023-1R KULSOM BA 세션



2023. 06. 02

고려대학교 오퍼레이션스 경영학회 KULSOM

Track A 논문팀 강지윤 권산 김규민 임상욱 정동은



목차

1. 기법 설명

1-1.인과추론 소개, 잠재적 결과 프레임워크

1-2.이중차분법 DID (Difference-in-Differences)

1-3.도구변수와 2SLS (Two-Stage Least Squares)

2. 논문 적용

2-1. 논문 소개 "Patient Routing to Skilled Nursing Facilities: The Consequences of the Medicare Reimbursement Rule"

2-2. Main question #1

2-3. Main question #3

01

기법 설명

1-1. 인과추론 소개, 잠재적 결과 프레임워크

◆ Causal Inference

- 인과추론(Causal Inference): 원인과 결과 사이의 인과관계를 추론하는 과정
- 인공지능 알고리즘은 데이터의 상관관계를 파악하여 분류, 예측 등의 문제를 해결한다. 하지만 상관관계는 인과관계를 의미하지 X
- 만약 인과관계를 추론할 수 있다면, 관찰된 상황과 다른 환경에서도 robust한 예측이 가능하고 복잡한 학습 모델의 특성을 이해할 수 있다.
→ Prediction을 주 목적으로 하는 Data Scientist와 인과 추론은 목적 자체가 다름

Indeed, they might play different roles with different goals.



선수의 성장에 주요 목적

Baseball Scouts

“He’s got a great swing. Natural swing.”

“Look, we’re gonna find 25 guys, put them through player development, teach them how to play Oakland A’s baseball.”



승리를 주요 목적

Data Scientists

“They’re still asking the wrong questions.”

“The goal shouldn’t be to buy players, what you want to buy is wins. To buy wins, you buy runs.”

◆ Potential Outcome Framework

- 인과추론을 위한 이상: "Treatment를 만약 받지 않았었다면 생기는 잠재적 결과(Counterfactual)"
- 인과추론의 현실: 실제로 Treatment를 받지 않은 Control group
→ 이상과 현실의 괴리감으로 인과추론이 어려움
- 이상적인 Counterfactual에 최대한 가까운 Control group과 비교를 해야 함
- Potential Outcome = Treatment를 했을 경우의 결과와 Treatment를 하지 않았을 경우의 잠재적 결과와의 차이



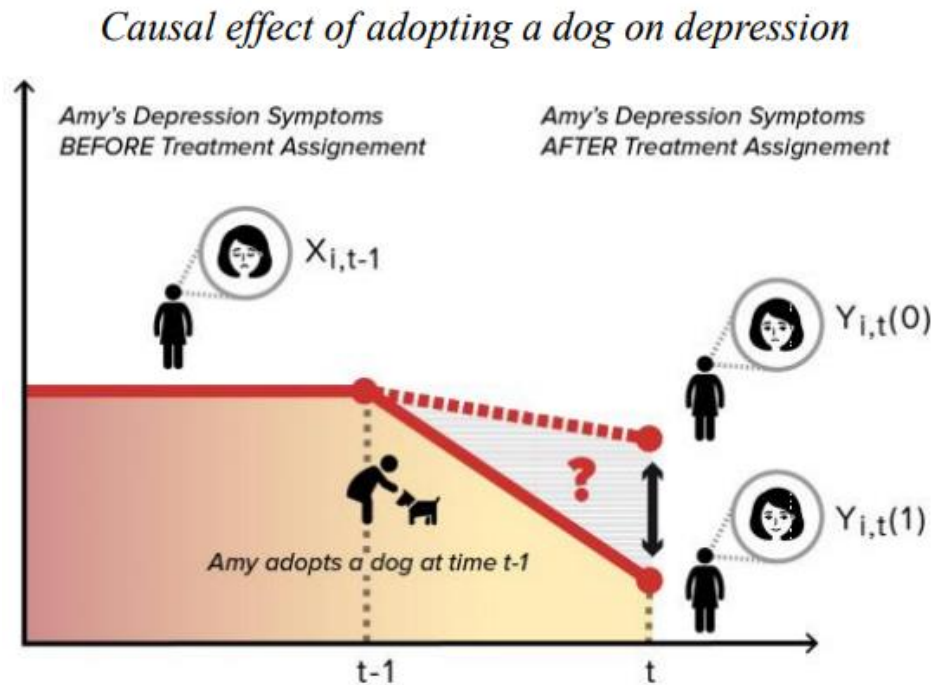
◆ Causal effect

- Causal effect: 이상 = (Actual outcome for treated if treated) – (Potential outcome for treated if not treated)
- Causal effect: 현실 = (Actual outcome for treated if treated) – (Actual outcome for untreated if not treated)

Counterfactual



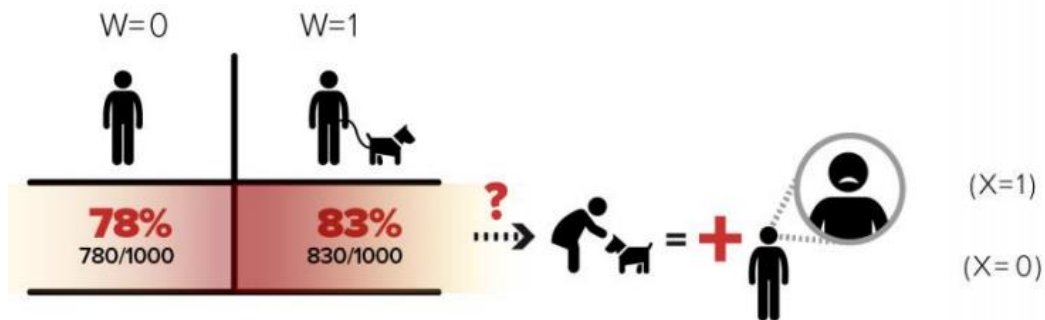
Control Group



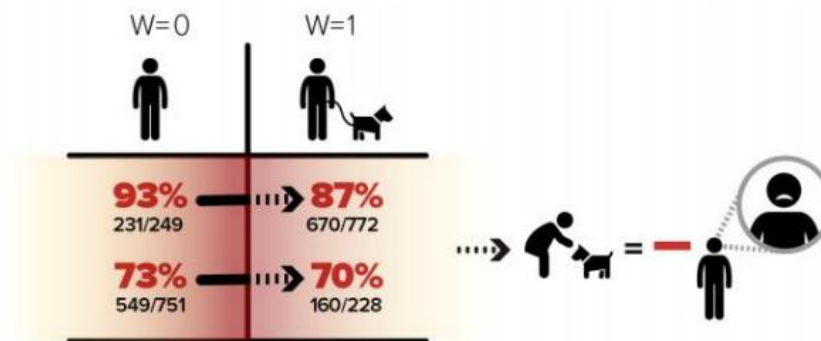
◆ Selection Bias

- 현실에서는 treatments가 random하게 할당되지 않음. 결과적으로 control group의 비교가 불가능할 수 있음
- 반려동물을 키우는 것은 우울증 감소에 도움이 될까?
- 연구결과 반려동물을 키우는 사람이, 반려동물을 키우지 않는 사람보다 우울증이 더 많음
→ 반려동물을 키워서 우울증에 걸리는 것이 아니라, 우울 성향이 높은 사람이 반려동물을 키우고자 하는 것을 수 있음

Are they comparable,
except the adoption W ?



Are they comparable,
except a confounder X and the adoption W ?



◆ Selection Bias

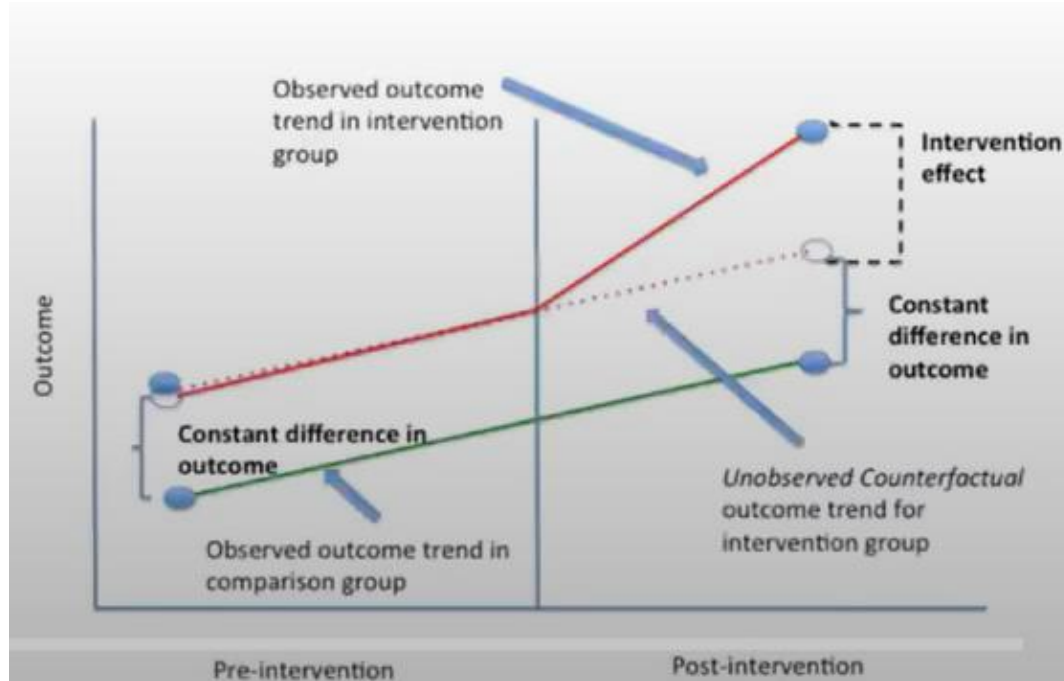
- 반려동물을 키우는 사람(Treatment) / 반려동물을 키우지 않는 사람(Control Group)
- 두 집단은 처음부터 다른 성향의 집단임. 즉 Control group, treatment group은 Treatment가 없더라도 서로 비교 불가
 - 두 집단을 비교하더라도 반려동물의 인과적 효과를 추정할 수 없다
 - 반려동물을 키우는 사실을 제외한 다른 요인들이 최대한 비슷한 집단으로 묶은 후, 비교해야 함
- $\text{Selection Bias} = \text{Control Group} - \text{Counterfactual}$
- 반려동물을 키울 요인을 통제함으로써, 차이를 '명시적으로 보정', 즉 Selection Bias를 보정

01

기법 설명

1-2. 이중차분법 DID (Difference-in-Differences)

What is DID?



- DID 는 말 그대로 차이의 차이를 의미
- 변인을 조정했을 때 그 효과를 측정하기 위함
- Treatment Group의 전후 차이 - Control Group의 전후 차이
- 변인을 조정하였을 때 조정하지 않은 상태에 대한 관측이 불가하기 때문에 이것에 대한 가정이 필요

What is DID?

	Before Treatment	After Treatment
Treatment Group	T_B	T_A (counterfactual: T'_A)
Control Group	C_B	C_A

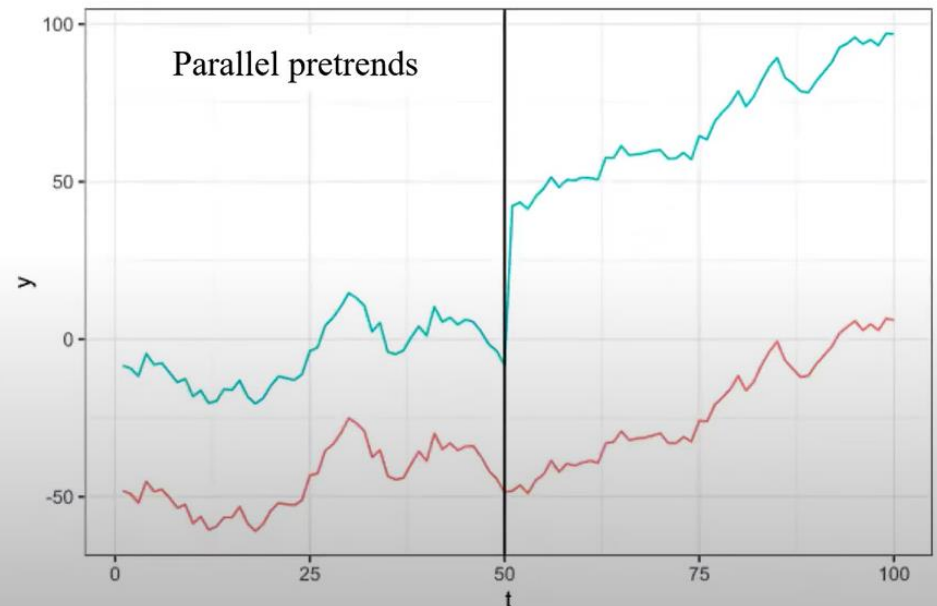
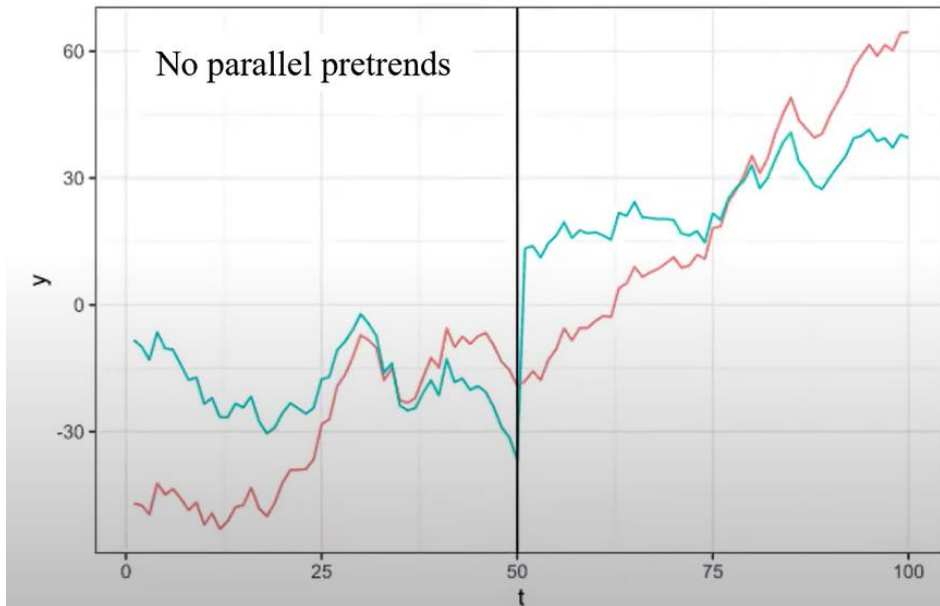
$$\begin{aligned} \text{DID estimator} &= (T_A - T_B) - (C_A - C_B) \\ &= T_A - [T_B + (C_A - C_B)] \\ &\quad \text{Inferred counterfactual} \end{aligned}$$

Counterfactual outcome in the absence of treatment	T_B	$T_B + (T'_A - T_B)$ Change for the treated in the absence of treatment (not observed)
Counterfactual outcome inferred from the control group	T_B	$T_B + (C_A - C_B)$ Inferred from the control group unaffected by the treatment

$$\text{Causal effect} = T_A - T'_A$$

Causal Effect = 실행으로 인한 효과 - 실행하지 않음으로 인한 효과

Parrel Trend Assumption



- DID 분석에 있어 Parallel Trend Assumption은 필수적
- Control Group 과 Treatment Group의 평행이 유지된다고 가정
- Treatment 전에 평행했던 것을 근거로 Treatment 후에도 평행하다고 추론

미국의 지역별 Uber 도입에 따른 교통사고 사망 사례의 변화

Table 6. Relative Time Model of Uber Entry on Alcohol Related Motor Vehicle Deaths

Dependent Variable	(1) ln(Num Deaths)	(2) ln(Num Deaths)
Model	Uber X	Uber Black
Rel Time _(t-4)	0.0428 (0.0280)	-0.0296 (0.0348)
Rel Time _(t-3)	-0.00251 (0.0270)	0.0116 (0.0361)
Rel Time _(t-2)	-0.0316 (0.0274)	-0.0138 (0.0362)
Rel Time _(t-1)	-0.0160 (0.0272)	0.00491 (0.0361)
Rel Time _(t0)	Omitted Base Case	
Rel Time _(t+1)	-0.0487* (0.0292)	-0.0154 (0.0346)
Rel Time _(t+2)	-0.0291 (0.0312)	0.0318 (0.0414)
Rel Time _(t+3)	-0.0530* (0.0314)	-0.0200 (0.0373)
Rel Time _(t+4)	-0.212*** (0.0705)	-0.0346 (0.0402)
Rel Time _(t+5)	-1.114*** (0.301)	-0.0270 (0.0390)
ln(Population)	-242.4 (665.4)	-34.69 (321.4)
ln(Median)	0.00978 (0.148)	0.0495 (0.145)
ln(Poverty)	-0.104 (0.0713)	-0.0939 (0.0658)
ln(Elderly)	0.122 (0.173)	0.128 (0.190)
ln(Police)	-0.00972 (0.0351)	-0.00628 (0.0306)
ln(College)	242.2 (665.5)	34.27 (321.6)
Constant	324.4 (885.1)	49.95 (425.9)
Time Fixed Effects	Yes	Yes
City Fixed Effects	Yes	Yes
N	12,420	12,420
R-squared	0.042	0.041

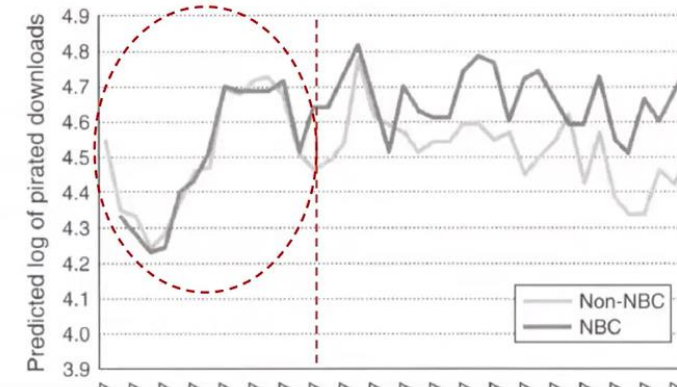
- 만약, Treatment가 동시에 일어나는 것이 아니라, 시간차를 두고 여러 번 일어난다면?
- 기존에는 Treatment 전을 0, 후를 1로 설정하여 계산을 진행함
- 해당 사례에서는 실행 년도에 따라 수치를 0~1사이로 보정하여 측정. 동시에 실행된 것이 아니라도 DID를 측정할 수 있음

기법 설명- 이중차분법 : DID (Difference-in-Differences)

Example : NBC's decision to remove its own content from iTunes



Figure 1 NBC vs. Non-NBC Piracy Surrounding December 1, 2007



- Treatment Group : NBC
- Control Group : Non-NBC
- Treatment 전에 NBC와 Non-NBC는 동일하게 진행하나, Treatment 이후 상이한 결과를 보이는 것으로 보아 Causal Effect 가 존재

기법 설명- 이중차분법 : DID (Difference-in-Differences)

Example : Effect of offline showrooms on demand generation and operation efficiency

FREE SHIPPING, HOME TRY-ONS & RETURNS SEE DETAILS CONTACT US HOME TRY-ON CART (0) SHOPPING CART (0)
LET'S FIND YOU SOME AMAZING GLASSES. (SIGN IN)

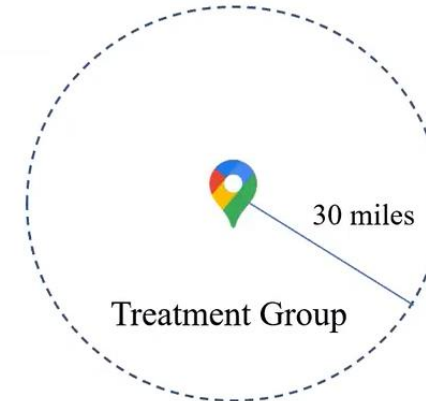
WARBY PARKER

MEN WOMEN BUY A PAIR, GIVE A PAIR OUR STORY MONOCLE HELP/FAQ



Warby Parker's website

Offline showroom (no purchase)

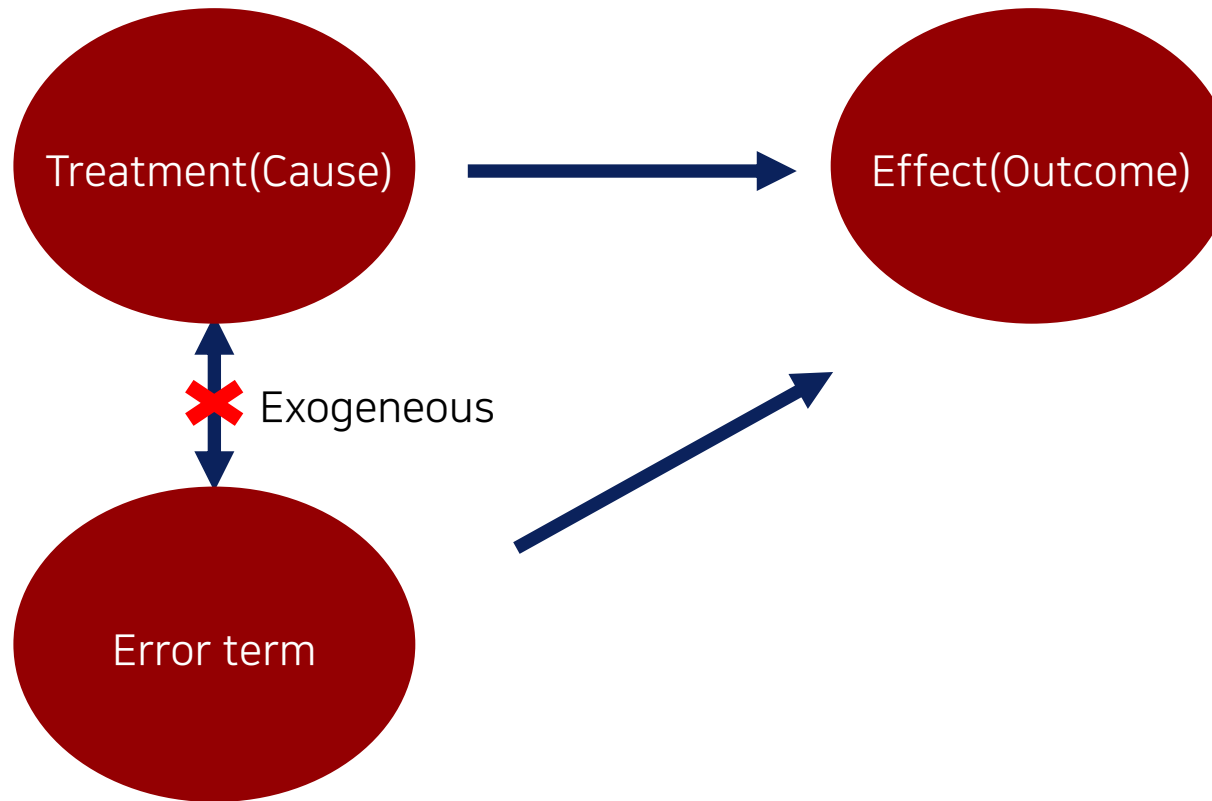


- Different treatment timing
- 시점이 다른 예시, 매장의 오픈 시점이 모두 다르다

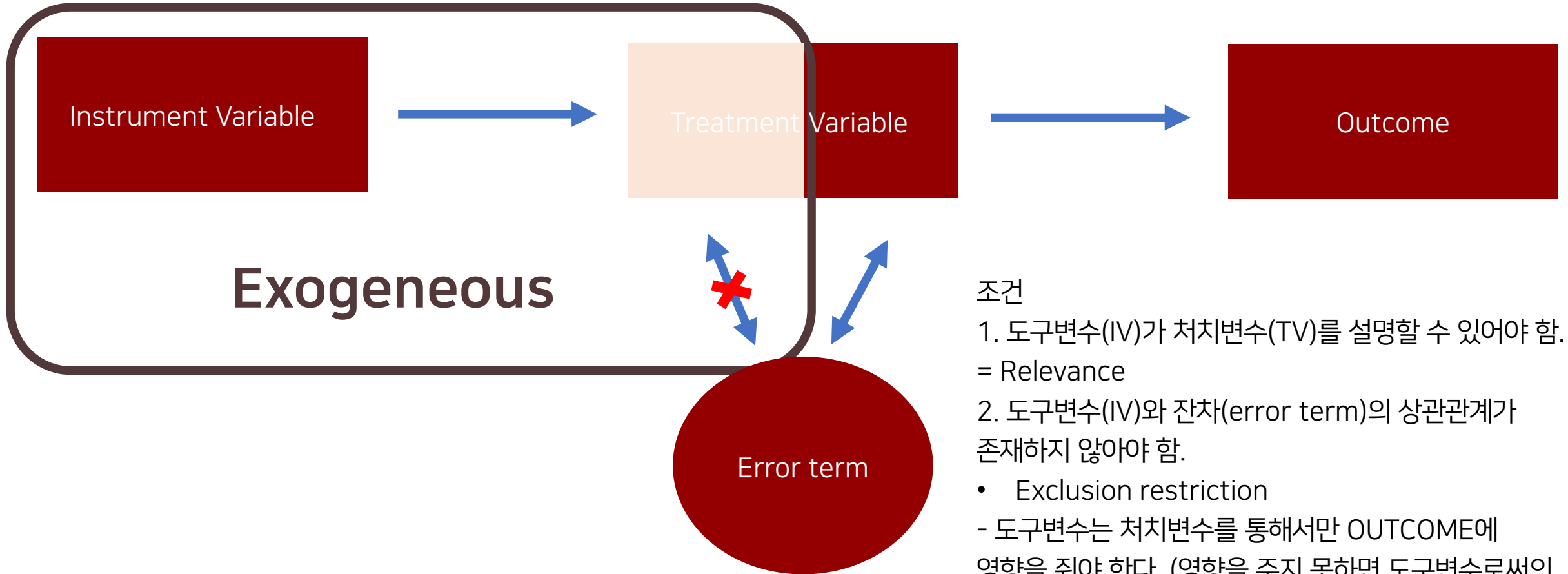
01

기법 설명

1-3. 도구변수와 2SLS (Two-Stage Least Squares)



- The difference b/w control group and counterfactual = selection bias, Unobservable selection bias = error term
 - $Cov(X, e | C) = 0$, 처치변수와 잔차 사이에는 어떠한 상관관계도 존재해서는 안된다. => 인과관계 해석 가능
- 그런데 만약 error term이 Treatment Variable과 endogeneous하더라도 100%의 상관관계를 갖는 것은 아님.
따라서, Treatment Variable을 endogeneous(selection bias)한 부분과 exogeneous(Causal relationship)한 부분으로 나눌 수 있음
Instrument Variable(IV)을 통해 통계적으로 둘을 구분해보자



조건

1. 도구변수(IV)가 처치변수(TV)를 설명할 수 있어야 함.
= Relevance
2. 도구변수(IV)와 잔차(error term)의 상관관계가 존재하지 않아야 함.
 - Exclusion restriction
 - 도구변수는 처치변수를 통해서만 OUTCOME에 영향을 줘야 한다. (영향을 주지 못하면 도구변수로서의 역할X)
 - Exogeneity of IV
 - 도구변수가 교란요인이 아니어야 함.

First stage : 도구변수를 통해서 exogeneous한 Treatment Variable만 예측

Second stage : endogeneous를 버린 Treatment Variable과 Outcome과의 인과관계 파악

1. Original equation

$$y = \beta_0 + \beta_1 X + \mu$$

2. First Stage

$$X = a_0 + a_1 z + v$$

$$a_1 \neq 0, Cov(z, \mu) = 0$$

도구변수 z로 설명할 수 있는 x(exogeneous), Residual = v (endogeneous)

3. Predicting the exogeneous portion of the treatment variable

$$\hat{x} = a_0 + a_1 z$$

4. Replacing x with \hat{x} in [1] for second stage

$$y = \beta_0 + \beta_1 \hat{x} + \mu$$

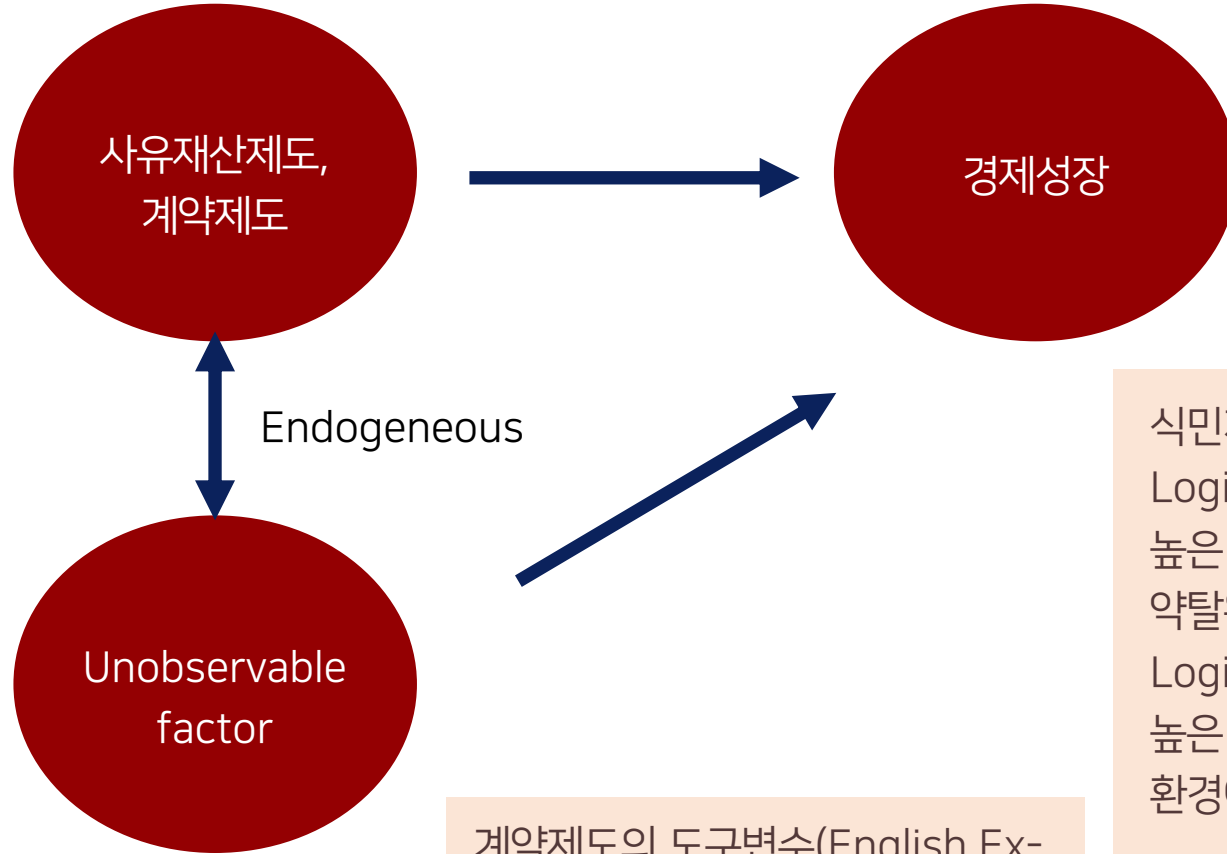
$$Cov(\hat{x}, \mu) = Cov(a_0 + a_1 z, \mu)$$

$$= a_0 Cov(1, \mu) + a_1 Cov(z, \mu) = 0$$

-> 인과관계로 해석가능

Example) 사회제도와 경제성장

사유재산제도, 계약제도(1990s) -> 경제성장(1990s)



참신한 도구변수 사용

“유럽제국주의 시절 침략국의 식민지에서의 사망률과 식민지에서의 인구 밀도”

식민지제도가 현대국가의 사회제도의 근간을 이룬다는 전제

Logic 1

높은 식민지 사망률(Due to 풍토병) -> 유럽국가 사회제도 도입x, 약탈위주(사유재산권 발달 x)

Logic 2

높은 인구 밀도 -> 발전, 자원풍부 -> 약탈 -> 사유재산권x
 환경이 유럽과 비슷했을 경우, 낮은 인구 밀도(자원x) -> 유럽 제도 도입

계약제도의 도구변수(English Ex-colonies)

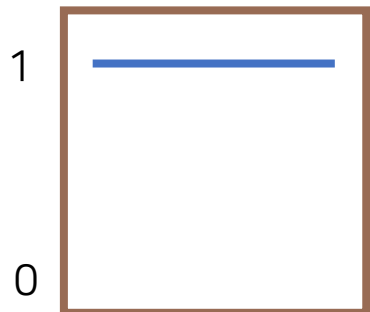
식민국가의 법 제도 : 영국의 영미법
 or 내륙국의 내륙법

멕시코와 미국

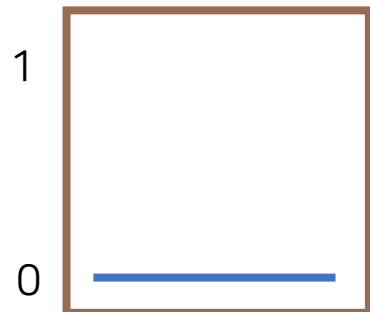
미국 = 허허벌판 / 영국 착취 x, 정착해서 삶, 사유재산제도 발달x

멕시코 = 발전문명 / 스페인에 착취, 사유재산제도 발달x

Potential Outcome Framework 관점에서의 2SLS, 2021 => 노벨 경제학상
 도구변수를 Treatment Assignment Mechanism에 적용, 2x2 matrix



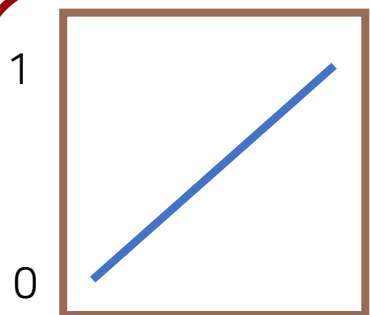
Always taker



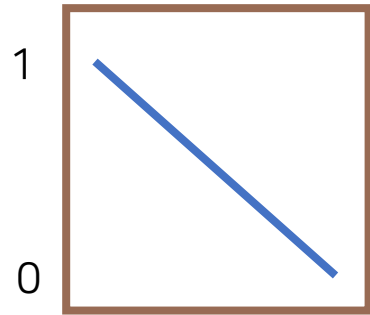
Never taker



도구변수에 의한 treatment 효과에 포함x



Complier



Defier

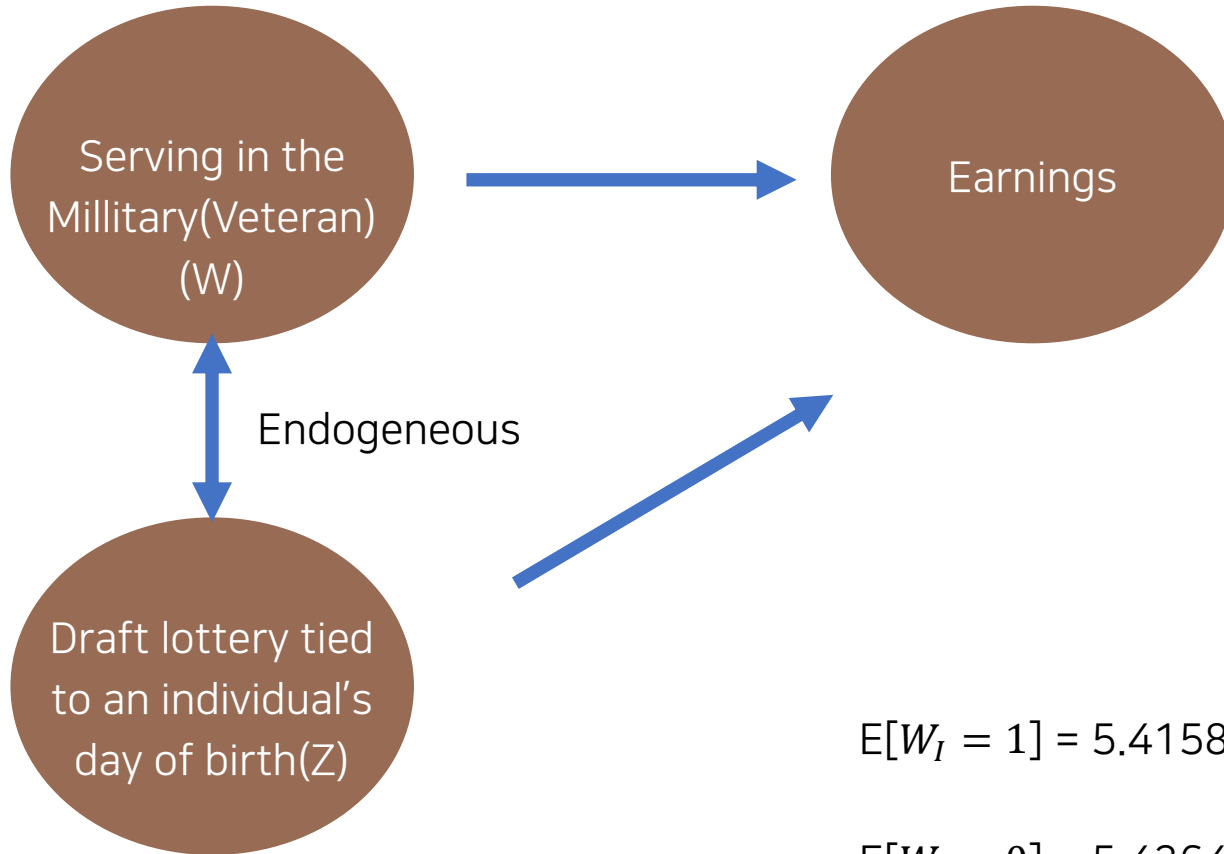


Sub-population에서의 인과관계 효과[LATE)

-도구변수에 의해 유도된 treatment의 효과

하지만, Defier의 경우 우리가 추정하고자 하는 인과관계 방해

Monotonicity assumption



Monotonicity assumption

높은 우선순위 -> 참전x

낮은 우선순위 -> 참전o

$$E[W_i = 1] = 5.4158$$

$$E[W_i = 0] = 5.4364$$

$$E[W_i = 1] - E[W_i = 0] = -0.0206$$

		# of participants Z_i	
		0	1
W_i	0	5,948	1,915
	1	1,372	865

		Expected Earnings Z_i	
		0	1
W_i	0	5.4472	5.4028
	1	5.4076	5.4289

OLS estimate

$$\text{Log}(\widehat{\text{earnings}})_i = 5.4364 - 0.0205\widehat{\text{Veteran}}$$

전쟁참여는 2% 손해? -> 전쟁참여자인 군인들에게
금전적인 지원을 해줄 근거x, 완벽한 인과관계x

2SLS estimate

$$\text{Log}(\widehat{\text{earnings}})_i = 5.4386 - 0.2336\widehat{\text{Veteran}}$$

$$E[W_i = 1 | Z_i = 1] - E[W_i = 0 | Z_i = 0] = -0.0183$$

우선 순위가 높을 때 참전하고 낮은 때 참전하지 않은
사람이라고 해석하면 안됨

어디서 나왔고 무엇을 의미하는지? 성인 남성 전체 혹은 특정한 사람?

Never taker의 비율은 0.6888

Always taker의 비율은 0.1874

Never taker와 always taker의 문제

W_i	Z_i	
	0	1
0	C / N	N / D
1	A / D	C / A

of participants

W_i	Z_i	
	0	1
0	5,948	1,915
1	1,372	865

W_i	Z_i	
	0	1
0	C / N	N / D
1	A / D	C / A

W_i	Z_i	
	0	1
0	0.6888 N	0 D
1	0.1237 C	0.1874 A

W_i	Z_i	
	0	1
0	5.4028	NA
$E(\widehat{Y_i}) = 0$		
1	5.6948	5.4076
$E(\widehat{Y_i}) = 1$		
5.4612		

전쟁참여로 평생 소득의 23% 손해 -> 모든 성인 남성에게 대한 효과x
 Draft Lottery에 의한 compliers의 인과적 효과

02

논문 적용

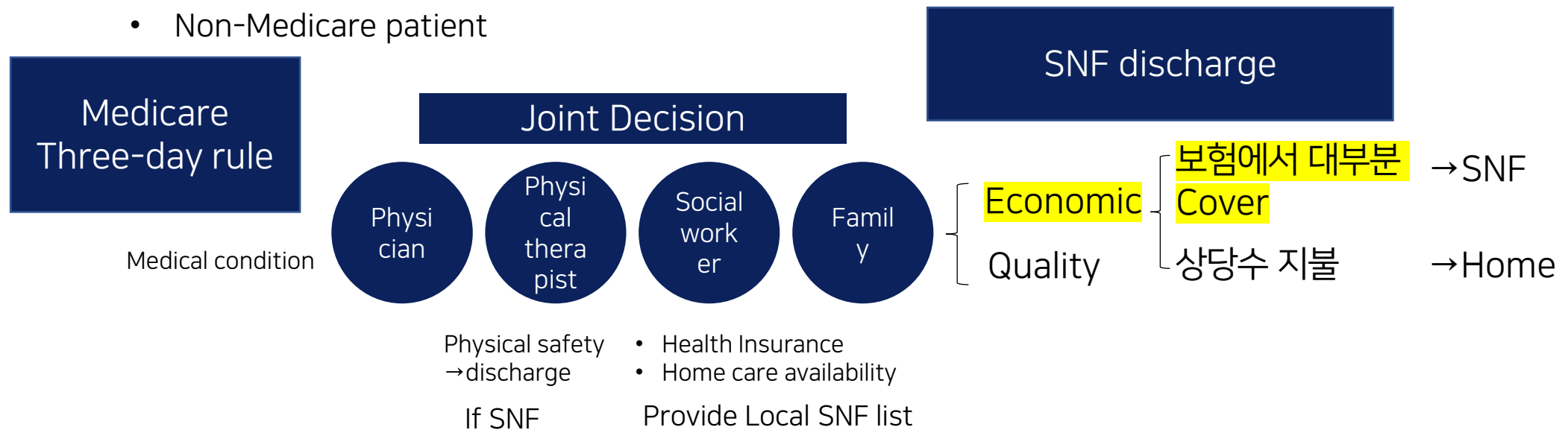
2-1. 논문 소개

◆ 핵심 내용 요약 (Introduction + Abstract)

- Three day rule: 최소 3일 입원한 사람 대상으로 skilled nursing care에 대해 첫 20일 full coverage 보상, 21-100일 partial coverage
 - SNF(skilled nursing facility)의 3일 규칙 → 사람들의 readmission(재입원)률 증가, overuse of SNF → Medicare 비용 증가
 - 결론: 사보험을 본 딴 머신러닝 알고리즘으로 SNF 3일 규칙을 대체해야 함.
-
- #Q1. How does the Medicare three-day rule affect the patient-routing decision in hospital discharge?
 - #Q2. Which type of postacute care, SNF care or home care, generates a better patient outcome, as measured by hospital readmission rate within 30 days?
 - #Q3. What alternative reimbursement rule should Medicare consider, if the current three-day-rule is proved inefficient?

◆ 키워드 정리

- Hospital discharge (SNF 전단계)
- Readmission
- SNF discharge (SNF로의 퇴원)
- Three-day rule = Medicare reimbursement rule
- Patient routing (SNF할지, home care할지)
- Postacute care (SNF)
- Medicare Day 2 patient
- Medicare Day 3 patient
- Non-Medicare patient



02

논문 적용

2-2. Main question #1: Hypothesis 1

Dataset: New York, Florida (2005–2015)의 600,000 이상 hospital inpatient discharge records

대상: 응급실에 접수된 Medicare (age 65–70) and non-Medicare (60–70) patients,
initially discharged from a hospital within 24 hours **before or after** the three-day cutoff.

제약 조건: Patient- and Hospital-level factors

Table 1. Summary Statistics by Length of Stay and Medicare Enrollment

Statistic	(1) Non-Medicare day 2	(2) Non-Medicare day 3	(3) Medicare day 2	(4) Medicare day 3
Discharge destinations				
SNF	0.009***	0.019***	0.017***	0.067
Home/home healthcare	0.929***	0.921***	0.921***	0.870
Other facilities	0.040***	0.045***	0.044***	0.051
Leaving against medical advice	0.022***	0.015***	0.019***	0.012
Patient characteristics				
Age	62.9***	62.9***	67.6*	67.6
Female	0.487***	0.491***	0.520***	0.528
White	0.577***	0.580***	0.650	0.651
Black	0.166***	0.171***	0.142***	0.148
Hispanic	0.155***	0.153***	0.141**	0.137
Comorbidities				
Any comorbidity	0.859***	0.878***	0.901***	0.918
Number of comorbidities	2.005***	2.218***	2.330***	2.592
Readmission rates				
30-day readmission, all-cause	0.073***	0.087***	0.088***	0.102
30-day readmission due to fall	0.002***	0.002***	0.002***	0.004
30-day readmission due to infection	0.016***	0.021***	0.021***	0.028***
30-day readmission due to avoidable condition	0.012***	0.015***	0.016***	0.020
Average occupancy rate	90.2***	90.2***	90.0	90.0
Average deficiency citations	5.6***	5.7***	5.8***	5.8
Average number of beds	177.8***	175.9***	166.7***	165.0
Observations	167,680	131,963	164,803	141,195

Notes. The sample consists of patients with (1) age between 60 and 70; (2) admitted through ER; (3) without prior hospitalizations in the past 180 days; and (4) treated by physicians with at least 100 records in New York and Florida between 2005 and 2015. Other facilities include various types of facilities, such as short-term hospitals, intermediate care facilities, long-term care hospitals, etc.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (for the difference between the respective category and the Medicare-day-3 group).

Group

Non-
Medicare
day2

Non-
Medicare
day 3

Medicare
day 2

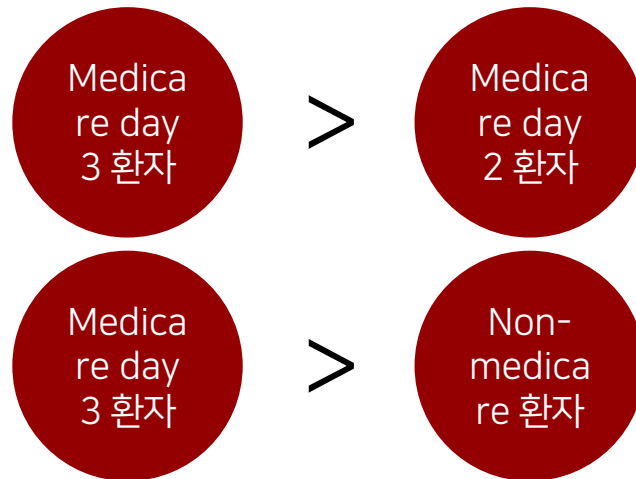
Medicare
day 3

Main question #1 How does the Medicare three-day rule affect the patient-routing decision in hospital discharge?

◆ Hypothesis 1 소개

- Hypothesis 1. Everything else being equal, the three-day rule should generate more SNF discharge for Medicare day 3 patients than Medicare day 2 patients. If non-Medicare insurance is less generous than Medicare and imposes more financial burden on the patient for SNF discharge, we should see a higher SNF discharge rate for Medicare day 3 patients than non-Medicare patients.*

SNF discharge rate



Main question #1 How does the Medicare three-day rule affect the patient-routing decision in hospital discharge?

◆ 1. DID의 적용 - Hypothesis 1

왜 DID? → Effect 크기 추정 (Three-day rule이 discharge destination에 미치는 영향 : Estimate effect of the three-day rule on discharge destination) *DID: 변인을 조정했을 때 그 효과를 측정하기 위함

$$Y_i = \alpha_0 + \alpha_1 \cdot Day3_i + \alpha_2 \cdot Medicare_i + \beta \cdot Day3_i \cdot Medicare_i + \delta \cdot X_i + e_i,$$

- i : 환자
 - $Day3_i$: Discharge after 24h
 - $Medicare_i$: 주요 기대 납부자(patients with Medicare)
 - 종속변수 Y_i :
 - (1) dummy variable
 - 1: SNF discharge 0
 - 0: SNF discharge X (including home/HHC, other facilities, or leaving against medical advice)
 - (2) dummy
 - 1: Home 0 (home or home healthcare)
 - 0: Home X (such as SNF, other facilities, or leaving against medical advice)
 - X_i : vector of discharge 특성들 (patients' demographics 등, hospital fixed effects)
- / 포함 0, 포함 X : robust standard errors / linear probability model + probit estimates for robustness

Medicare (age 65-70)/non-Medicare (age 60-70)환자의 SNF discharge 여부
Control on: 병원에서 3일 기한 24h Before vs After

- β : Medicare day 3, day 2 patients' discharge difference relative to difference among non-Medicare day 3, day 2 patients *conditional on patient demographics, health conditions, seasonality, and discharge hour.
- $\beta \rightarrow$ 가설 1이 맞을 경우, significantly positive

DID의 전제 : Parallel trend assumption

: Medicare와 non-Medicare 치료 없을 때의 결과 사이
→ 검증: pretrend를 raw data에서 검사 (뒷 슬라이드)

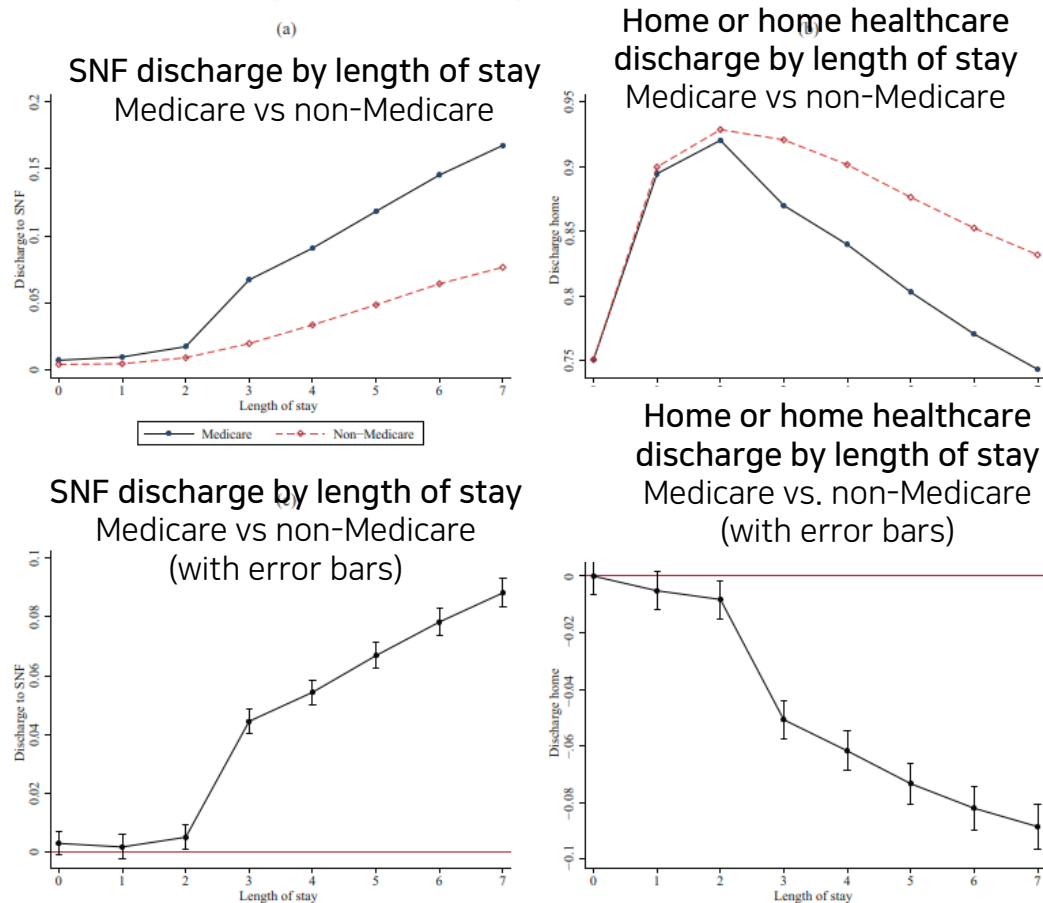
추가 검증(Alternative): Medicare eligibility(65세)에 대한 Regression discontinuity design

→ Medicare eligibility가 three-day rule의 discharge destination에 영향을 주는지 : 65세의 discharge destination에 대한 day 2와 day3의 변화 비교

The Overall Effect of the Three-Day Rule on SNF Discharge

- DID를 적용하여 three-day rule이 discharge destination에 미치는 영향을 측정
- Treatment group: Medicare
- Control group: non-Medicare

Figure 1. (Color online) Examining Parallel Trends in Discharge Destination



Notes. (a) SNF discharge by length of stay, Medicare vs. non-Medicare. (b) Home or home healthcare discharge by length of stay, Medicare vs. non-Medicare. (c) SNF discharge by length of stay, Medicare vs. non-Medicare, with error bars. (d) Home or home healthcare discharge by length of stay, Medicare vs. non-Medicare, with error bars.

*mean discharge rate to SNF **coefficients on the interactions between each length of stay and Medicare from a regression of a discharge destination / Medicare:1, non: 0

- Examine raw data → Find strong evidence (우리가 설정한 Medicare와 non-Medicare 환자의 parallel trend가 plausible하다)

(a) SNF: 0-2일까지 비슷하다가 3일부터 Medicare의 큰 폭 증가
 (b) Home: 0-2일까지 비슷하다가 3일부터 Medicare의 큰 폭 감소
 (c), (d): Coefficients(length of stay&medicare): 0-2일까지 비슷하다가 3일부터 큰 폭으로 discharge destination이 갈림.

Medicare
day 3
discharged

SNF ↑
Home ↓

논문 적용 - (1) DID Analysis

The Overall Effect of the Three-Day Rule on SNF Discharge

Without control 변수

With control 변수

Table 2. The Effect of the Three-Day Rule on Discharge Destinations

Dependent variable	(1)	(2)	(3)	(4)
	Without control variables		With control variables	
	SNF	Home/home healthcare	SNF	Home/home healthcare
<i>Day 3 × Medicare</i>	0.039*** (0.002)	-0.043*** (0.003)	0.034*** (0.002)	-0.034*** (0.002)
Observations	605,641	605,641	605,641	605,641
Mean Y	0.027	0.911	0.027	0.911

Notes. This table reports regression coefficients (β) from Equation (1). Columns (1) and (2) show the estimates without the inclusion of control variables. Columns (3) and (4) show the results including control variables, such as a series of indicators for age, sex, race, diagnosis-related group fixed effects, indicators for all comorbidities, an indicator for hospital bed availability, hospital fixed effects, year fixed effects, month fixed effects, and discharge hour dummies. Robust standard errors, clustered by health service area, are in parentheses.

*** $p < 0.001$.

DID 추정 결과

Treatment group: Medicare patients

Control group: non-Medicare patients

의 day2와 day3의 discharge rate 변화 차이

=Effect 비교

Medicare day 3
discharged

SNF ↑, Home ↓

Without control 변수

Three-day rule은 SNF로의 discharge를 3.9%p 증가시키고,
home으로의 discharge를 4.3%p를 감소

With control 변수

Control 변수를 포함함으로써, 추정 결과(estimate)의 크기가 줄음.
Three-day rule이 SNF discharge를 3.4%p 증가,
Home discharge를 3.4%p 감소

Three-day rule의
SNF로의 Discharge 증가 효과가
Home으로의 Discharge 감소
효과에 의해 상쇄됨(Offset)

1. Effect Size가 큰 것인가? 2. Medicare Three-day rule의 영향이 맞는가?

추정 Discharge 확률 (by. 입원 기간/Medicare 등록 여부)

Table B.3: The estimated discharge probabilities by length of stay and Medicare enrollment

4 groups	Without control 변수		With control 변수	
	Without control variables		With control variables	
	Day 2	Day 3	Day 2	Day 3
	Panel A. Dependent variable: discharge to SNF			
Non-Medicare	0.009	0.019	0.011	0.019
Non-Medicare SNF	[0.007, 0.011]	[0.017, 0.022]	[0.006, 0.017]	[0.013, 0.024]
Medicare	0.017	0.067	0.019	0.062
Medicare SNF	[0.015, 0.019]	[0.061, 0.073]	[0.013, 0.024]	[0.055, 0.069]
Panel B. Dependent variable: discharge home or HHC				
Non-Medicare	0.929	0.921	0.926	0.923
Non-Medicare Home	[0.921, 0.936]	[0.914, 0.928]	[0.916, 0.935]	[0.913, 0.932]
Medicare	0.921	0.870	0.917	0.877
Medicare Home	[0.914, 0.927]	[0.862, 0.878]	[0.907, 0.926]	[0.866, 0.887]

Notes: This table reports the estimated discharge probabilities to each destination separately for each of the four groups. The corresponding 95% confidence intervals are in brackets. Controls include a series of indicators for age, sex, race, diagnosis related group fixed effects, indicators for all comorbidities, an indicator for hospital bed availability, hospital fixed effects, year fixed effects, month fixed effects, and discharge hour dummies.

Three-day rule의 SNF discharge 효과는 거대함

- Medicare day 2 → SNF : only 1.9%
- Medicare day 3 → SNF : 6.2%

Non-Medicare → SNF: day2(1.1%), day3(1.9%)보다 차이가 큼.

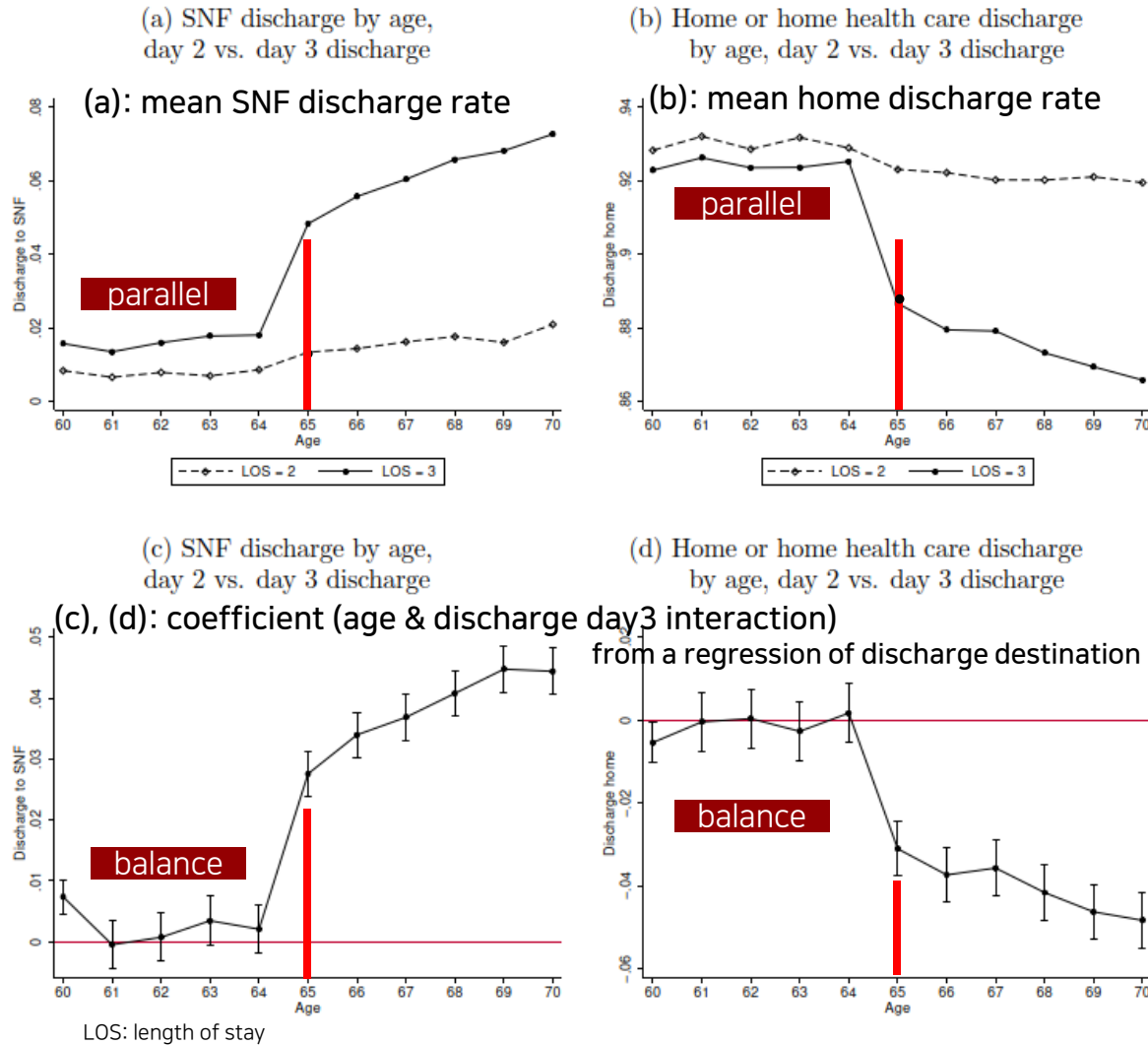
- Medicare day 2 → Home : 91.7%
- Medicare day 3 → Home : 87.7%

Non-Medicare → Home: day2(92.6%), day3(92.3%)보다 차이가 큼.

Raw data 결과(Table 1)와 합쳤을 때,
4 그룹 SNF discharge rate의 차이
(raw difference) 중 90%는
Medicare Three-day rule로 설명 가능하다

Medicare Eligibility의 영향: 65세 자격 규정 / RD(Medicare eligibility)

Figure A.1: Examining parallel trends in discharge destination



Regression Discontinuity

Regression discontinuity를 적용
(Medicare 자격 규정: 65세)

Discharge probability가 age에 따라 어떻게 변하는지
 (Day 2 vs Day 3 비교)

(a), (b): 60-64세까지 day 2, day 3 모두 parallel trend

But 65세 day 2와 day 3 큰 divergence

(c), (d): 65세 이전에는 discharge destination이
 balanced (coefficient 0에 가까움).

But 65세 이후 day 3 discharges 중에서

SNF(home) discharge rate: 큰 증가(감소)폭

Regression Discontinuity

Table B.4: The effect of the three-day rule on discharge destinations, regression discontinuity design approach based on patient age

	SNF	Home/Home health care
Above 65	-0.005 (0.005)	-0.002 (0.008)
Above 65 × Day 3	0.021*** (0.002)	-0.024*** (0.004)
Observations	451093	451093
Mean Y	0.027	0.911

Notes: This table reports regression coefficients (γ_1 and γ_2) from the following equation:

$$\begin{aligned}
 Y_i = & \delta_0 + \gamma_1 1[Age_i \geq 65] + \delta_1 (Age_i - 65) + \delta_2 1[Age_i \geq 65] \times (Age_i - 65) \\
 & + \gamma_2 1[Age_i \geq 65] \times Day3_i + \delta_3 (Age_i - 65) \times Day3_i \\
 & + \delta_4 1[Age_i \geq 65] \times (Age_i - 65) \times Day3_i + \delta_5 Day3_i + \delta_6 X_i + e_i
 \end{aligned}$$

Controls (X_i) include a series of indicators for age, sex, race, diagnosis related group fixed effects, indicators for all comorbidities, an indicator for hospital bed availability, hospital fixed effects, year fixed effects, month fixed effects, and discharge hour dummies. Robust standard errors, clustered by health service area, are in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

: day 2와 day 3의
65세에서의 불연속의 크기에 대한 차이
(coefficient)

Medicare eligibility (65세 자격 규정)

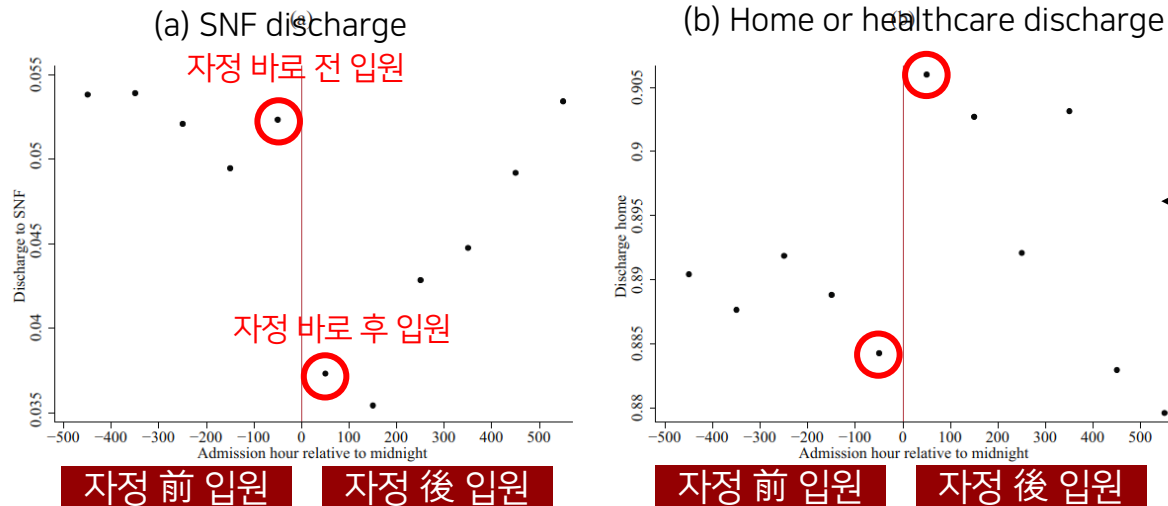
→ day 3 discharge 中 SNF 크게 증가

→ day 3 discharge 中 Home 크게 감소
(day 2 discharge에서는 little impact)

Three-day rule이 SNF discharge를
significantly하게 증가시킴.

Three-day rule의 discharge destination에 대한 영향 - 입원 시간 / RD(admission hour)

Figure 2. (Color online) Discharge Rates by Admission Hour Relative to Midnight



Regression Discontinuity

RD (admission hour)

검증: 3일 기한의 경과가 discharge destination에 미치는 영향
한정: 1. 자정(midnight)의 5시간 전후의 입원 건에 대해서만 집중
2. Sample을 한정 (Hospital에서 60-84시간 안에 퇴원한 환자)

Figure 2: 자정 바로 전에 입원 -> Three-day rule을 더 쉽게 충족
-> SNF로 더 많이 discharge/home으로 더 적게
(자정 바로 뒤에 입원한 환자보다)

Table B.5: The effect of the three-day rule on discharge destinations, regression discontinuity design based on admission around midnight

Dependent variable	(3) With control variables			
	SNF	Home/home healthcare	(1) SNF	(2) Home/Health care
Day 3 × Medicare	0.034*** (0.002)	-0.034*** (0.002)	0.015*** (0.004)	-0.021** (0.006)
Observations	605,641	605,641	50302	50302
Mean Y	0.027	0.911	0.049	0.892

↑
앞선 결과에서 나왔던
DID 분석에서의 추정치

Notes: The sample is restricted to patients who are discharged within 60-84 hours from admission. This table reports a regression coefficient (λ_1) from the following equation:

$$Y_i = \lambda_0 + \lambda_1[AdmissionHour_i < 2400] + \lambda_2(AdmissionHour_i - 2400) + \lambda_31[AdmissionHour_i < 2400] \times (AdmissionHour_i - 2400) + e_i$$

Robust standard errors, clustered by health service area, are in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table B.5: RD(admission hour)에서의 coefficients를 나타냄.
→ Three-day rule이 추가적으로 SNF로의 discharge를 생성했다는 것을 검증
*DID에서의 추정치보다 다 크기가 작은데, RD로 midnight 근방의 좁은 범위만 다뤘기 때문.

논문 적용 - (2) Regression Discontinuity Design

Three-day rule의 discharge destination에 대한 영향 - 입원 시간 / RD(admission hour)

Table B.6: The effect of the three-day rule on discharge destinations, robustness checks

	(1) SNF	(2) Home/Home health care
Panel A. Dropping Medicaid patients		
Day 3 × Medicare	0.036*** (0.002)	-0.035*** (0.002)
Observations	534847	534847
Mean Y	0.026	0.914
Panel B. Dropping Medicare Advantage enrollees		
Day 3 × Medicare	0.047*** (0.002)	-0.047*** (0.003)
Observations	489816	489816
Mean Y	0.027	0.908
Panel C. Dropping years 2012 and after		
Day 3 × Medicare	0.030*** (0.002)	-0.031*** (0.003)
Observations	326869	326869
Mean Y	0.025	0.908
Panel D. Restricting to ages 62 to 67		
Day 3 × Medicare	0.030*** (0.002)	-0.030*** (0.003)
Observations	323528	323528
Mean Y	0.024	0.914

	(1) SNF	(2) Home/Home health care
Panel E. Dropping 60-64 non-Medicare patients		
Day 3 × Medicare	0.026*** (0.002)	-0.027*** (0.003)
Observations	362346	362346
Mean Y	0.037	0.900
Panel F. Medicare and privately insured patients only		
Day 3 × Medicare	0.035*** (0.002)	-0.033*** (0.002)
Observations	475275	475275
Mean Y	0.028	0.914
Panel G. Restricting to 20 most common DRG codes with SNF discharge		
Day 3 × Medicare	0.061*** (0.004)	-0.063*** (0.004)
Observations	152252	152252
Mean Y	0.044	0.901
Panel H. Including physicians with at least 50 records		
Day 3 × Medicare	0.034*** (0.002)	-0.035*** (0.002)
Observations	745970	745970
Mean Y	0.027	0.927
Panel I. Using probit models		
Day 3 × Medicare	0.015*** (0.001)	-0.028*** (0.002)
Observations	593147	602764
Mean Y	0.027	0.911

*Medicaid: 미국의 저소득층 의료 보장 제도

*Medicare Advantage: 미국의 건강우대보험

Robustness 분석

→ 결과가 robust한지?

Alternative sample 제한 조건에 대하여

A. control group에서 저소득층 의료 보장 환자 제외

B. control group에서 건강우대보험 환자 제외

C. 시기: 재입원 축소 정책 적용 후 시기 제외

D. 연령: 연령대를 62-67로 줄이기

E. 연령: 60-64세의 non-Medicare 환자 제외

F. 환자: Medicare과 사보험 가입 환자만 집중

G. SNF에 흔한 진단만 집중(diagnoses common for SNF)

H. 최소 50 건 이상의 physician만 포함시키기

I. Robustness 검증

: OLS(최소자승모형, ordinary least squares)

대신 Probit 활용

결론: 가설 1을 뒷받침한다.


◆ Main question #1 - Hypothesis 1 :

How does the Medicare three-day rule affect the patient-routing decision in hospital discharge?

→ DID analysis + Regression Discontinuity(RD) 적용

3일 기한(Three-day cutoff) After 0-24h 경과 환자

3일 기한(Three-day cutoff) Before 0-24h 경과 환자



After: Before보다 3.4%p 더
SNF Discharge

*이 연구의 sample에서는 고작 4%의 Medicare 환자 (+1.3%의 non-Medicare 환자)가 SNF로 discharge됨. = enormous effect

**Regression Discontinuity(RD) approach를
admission hour에 기반해 적용 시 → 결과가 Robust

02

논문 적용

2-3. Main question #3

가설: SNF discharge(SNF로의 퇴원)은 30일 병원 재입원과 인과관계가 있을 것이다

SNF로의 퇴원



30일 병원 재입원
(건강)

◆ 'SNF로의 퇴원'은 내생적이기 때문에 인과 관계의 파악이 어려움

- 환자들은 다양한 이유로 'SNF로 퇴원'
 - 그리고 그 이유가 결과(건강)에도 동시에 영향을 미침
 - EX1) 치료가 필요한 환자(건강이 안 좋은 환자)는 SNF에서의 후기 치료를 원할 수 있음
 - EX2) 의사는 환자들이 SNF에서 더 나은 치료를 받도록, SNF로 퇴원시킬 수 있음
- 'SNF로의 퇴원' 과 '30일 병원 재입원' 간의 정확한 인과관계의 파악이 어려워짐

◆ 내생성을 배제하고자 2SLS의 사용

- 두 변수 간의 정확한 인과관계를 추론하고자 한다면
- 내생성을 배제하고 외생성을 확보해야 함
- Instrumental Variable(도구 변수)의 도입을 통해, 'SNF로의 퇴원(독립 변수)'에서 외생적인 부분만 고려하여 인과관계 파악

Instrument Variable



SNF로의 퇴원
(외생적) (내생적)



30일 병원 재입원(건강)

가설의 2sls 공식에의 적용(도구변수로 Physicianm의 사용)



First stage : 도구변수를 통해서
exogeneous한 Treatment Variable만 예측

$$SNF_i = \mu + \gamma \cdot Physician_m + \rho \cdot X_i + \varepsilon_i,$$

- SNFi: 환자가 SNF로 퇴원하는 경우를 나타내는 이진 변수
- μ : 회귀식의 절편(모델의 기준점)
- γ : Physicianm 변수의 회귀 계수(coefficient)
- Physician: 비슷한 상태의 환자들을 조금 더 SNF로 퇴원시키는 경향
- ρ : 설명 변수 X_i 에 대한 회귀 계수(coefficient)
- ~ 설명 변수 X_i 가 종속 변수에 어떤 영향을 미치는지(크기와 방향)
- X_i : 다양한 퇴원 특성(환자의 연령, 성별, 질병 상태 등)
- ε_i : 오차항

Second stage : endogeneous를 버린
Treatment Variable과 Outcome과의 인과관계 파악

$$Readmission_i = \lambda + \theta \cdot \widehat{SNF}_i + \pi X_i + \eta_i,$$

- Readmission_i: 병원 재입원 여부를 나타내는 지표
- λ : 회귀식의 절편(모델의 기준점)
- π : 설명 변수 X_i 에 대한 회귀 계수(coefficient)
- ~ 독립 변수 X_i 종속 변수에 어떤 영향을 미치는지(크기와 방향)
- θ : 'SNF로의 퇴원' 이 '30일 병원 재입원'에 미치는 영향
- SNFi: 첫 번째 단계에서 추정된 SNF 퇴원의 예측값
- π : 설명 변수 X_i 의 회귀 계수
- X_i : 추가적인 설명 변수
- η_i : 오차 항

예시를 통한 이해: OSL 방식을 바탕으로 잔차를 최소화하는 계수 계산

First stage : 도구변수를 통해서
exogeneous한 Treatment Variable만 예측

첫번째 수식

$$SNF_i = \mu + \gamma \cdot Physician_m + \rho \cdot X_i + \varepsilon_i,$$

$$SNF_i = \beta_0 + \beta_1 \cdot Physician_m + \beta_2 \cdot X_i + \varepsilon_1$$

Physician_m: [0.5, 0.7, 0.3, 0.8, 0.6]

SNF_i: [1, 0, 1, 0, 1]

X_i: [2, 1, 3, 2, 2]

$$SNF_1 = \beta_0 + \beta_1 \cdot 0.5 + \beta_2 \cdot 2 + \varepsilon_1$$

$$SNF_2 = \beta_0 + \beta_1 \cdot 0.7 + \beta_2 \cdot 1 + \varepsilon_1$$

$$SNF_3 = \beta_0 + \beta_1 \cdot 0.3 + \beta_2 \cdot 3 + \varepsilon_1$$

$$SNF_4 = \beta_0 + \beta_1 \cdot 0.8 + \beta_2 \cdot 2 + \varepsilon_1$$

$$SNF_5 = \beta_0 + \beta_1 \cdot 0.6 + \beta_2 \cdot 2 + \varepsilon_1$$

- 최소자승법(OLS)을 사용하여 'ε₁의 제곱합'을 최소화하는 회귀 계수를 추정

→ 각각의 SNF 값에 대한 추정값(내생성이 배제된 값)을 계산

Second stage : endogeneous를 버린
Treatment Variable과 Outcome과의 인과관계 파악

두번째 수식

$$Readmission_i = \lambda + \theta \cdot \widehat{SNF}_i + \pi X_i + \eta_i,$$

$$Readmission_i = \gamma_0 + \gamma_1 \cdot SNF_i + \gamma_2 \cdot X_i + \varepsilon_2$$

SNF_i: [0.7, 0.3, 0.8, 0.5, 0.6]

X_i: [2, 1, 3, 2, 2]

Readmission_i: [0, 1, 1, 0, 0]

$$Readmission_1 = \gamma_0 + \gamma_1 \cdot 0.7 + \gamma_2 \cdot 2 + \varepsilon_2$$

$$Readmission_2 = \gamma_0 + \gamma_1 \cdot 0.3 + \gamma_2 \cdot 1 + \varepsilon_2$$

$$Readmission_3 = \gamma_0 + \gamma_1 \cdot 0.8 + \gamma_2 \cdot 3 + \varepsilon_2$$

$$Readmission_4 = \gamma_0 + \gamma_1 \cdot 0.5 + \gamma_2 \cdot 2 + \varepsilon_2$$

$$Readmission_5 = \gamma_0 + \gamma_1 \cdot 0.6 + \gamma_2 \cdot 2 + \varepsilon_2$$

- 최소자승법(OLS)을 사용하여 'ε₂의 제곱합'을 최소화하는 회귀 계수를 추정

→ 'SNF로의 퇴원'의 '30일 병원 재입원'에 대한 영향 θ 계산

도구변수의 조건

1. 도구변수(IV)가 처치변수(TV)를 설명할 수 있어야 함.

가정1. 도구변수(Physician_m)가 종속변수(SNFi)와 상관관계를 가지는 것을 가정

도구변수의 조건

2. 도구변수(IV)와 잔차(error term)의 상관관계가 존재하지 않아야 함.

가정2. 도구변수(Physician_m)가 오차항(η_i)과 상관관계가 없다는 것을 가정

$$SNF_i = \mu + \gamma \cdot Physician_m + \rho \cdot X_i + \varepsilon_i,$$

Table B.9: First stage: The relationship between the instrument and SNF discharges

	(1) Non-Medicare day 2	(2) Non-Medicare day 3	(3) Medicare day 2	(4) Medicare day 3
Panel A. OLS				
Doctor's tendency to use SNF	0.550*** (0.127)	0.712*** (0.101)	0.632*** (0.073)	1.085*** (0.031)
Observations	157146	124026	154455	132267
Mean Y	0.009	0.021	0.018	0.071

가정1. 도구변수(Physician_m)가 종속변수(SNF_i)와 상관관계를 가지는 것을 가정

모든 그룹에서 $p < 0.001$: 도구변수(Physician_m)와 종속변수(SNF_i)의 관계가 통계적으로 유의함

→ 이는 가정이 부합하는 것을 시사

$$Readmission_i = \lambda + \theta \cdot \widehat{SNFi} + \pi X_i + \eta_i,$$

Table 3. The Effect of SNF Discharge on Hospital Readmission Within 30 Days, 2SLS

Estimation group	(1) Non-Medicare day 2	(2) Non-Medicare day 3	(3) Medicare day 2	(4) Medicare day 3
Panel A. All-cause readmission				
Discharge to SNF	-0.013 (0.038)	0.025 (0.036)	0.053 (0.051)	0.090** (0.027)
Observations	157,146	124,026	154,455	132,267
Mean Y	0.073	0.087	0.088	0.104
First-stage F-statistic	18.814	49.316	75.046	1,244.143
Panel C. Readmission due to infection				
Discharge to SNF	0.016 (0.016)	0.015 (0.016)	0.040+ (0.023)	0.055*** (0.012)
Observations	157,146	124,026	154,455	132,267
Mean Y	0.016	0.021	0.021	0.029
First-stage F-statistic	18.814	49.316	75.046	1,244.143

- 추가적으로, 다른 재입원 원인을 조사한 결과, Medicare day 3 환자들의 재입원 증가는 감염 관련 진단으로 확인됨
- Medicare day 3 환자들 그룹(3일 규칙에 해당되는 유일한 그룹)에서 'SNF 퇴원'이 '30일 이내의 재입원(감염 관련 재입원)'을 유의하게 5.5%p 증가시킴
- 3일규칙은 '30일 병원 재입원(감염 관련 재입원)'을 0.2%p증가시킴(5.5×0.055)

- Medicare day 3 환자들 그룹(3일 규칙에 해당되는 유일한 그룹)에서 'SNF 퇴원'이 '30일 이내의 재입원(모든 원인에 의한 재입원)'을 유의하게 9%p 증가
- 3일 규칙으로 인해 SNF 퇴원율이 3.4%p 증가하는 것을 감안할 때, 이는 '3일 규칙'이 '30일 재입원율'을 0.3%p(3.4×0.09) 증가한다는 것을 의미함

- ◆ Robustness Test(강건성 평가): 다양한 상황에서의 결과의 일관성과 안정성을 평가하는 것을 의미
→ 연구 결과의 신뢰성 및 일반화 가능성 보장
- 민감도 분석(Sensitivity analysis): 연구 결과에 영향을 미치는 주요 변수들을 변동시켜가면서 결과의 변동을 관찰

$$Readmission_i = \lambda + \theta \cdot \widehat{SNFi} + \pi X_i + \eta_i$$

$$Readmission_i = \lambda + \theta \cdot SNFi + \pi_1 \cdot Physician_m + \pi_2 \cdot X_i + \eta_i$$

	(1) 2SLS	(2) 2SLS with a readmission control
Panel A. Specification check		
Discharge to SNF	0.055*** (0.012)	0.036** (0.011)
Observations	132267	132267
Mean Y	0.029	
First stage F statistic	1244.143	1208.836

- ◆ 가정2. 도구변수(Physician_m) 가 오차항(η_i)과 상관관계가 없다는 것을 가정

- 도구와 오차항(η_i) 사이에 상관관계 0: 만약 도구 변수(Physician_m)의 계수가 유의미한 결과를 나타내지 않고 도구를 포함하여 추정결과가 크게 변하는 경우
- 도구와 오차항(η_i) 사이에 상관관계 X: 도구를 포함하더라도 추정 결과가 크게 변하지 않는 경우

- ◆ 결과 분석

- θ : 0.055 → 0.036
- 두 추정치의 차이가 통계적으로 유의미한 차이가 아니다

→ 가정2의 검증 완료

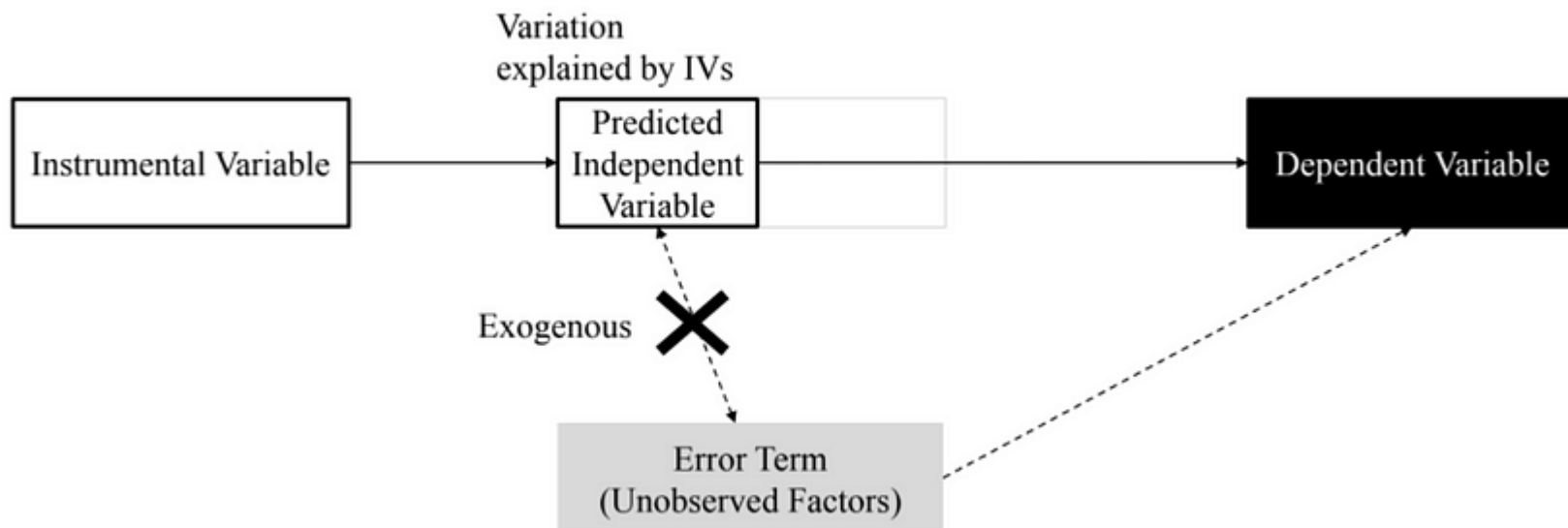
- ◆ Robustness Test(강건성 평가): 다양한 상황에서의 결과의 일관성과 안정성을 평가하는 것을 의미
→ 연구 결과의 신뢰성 및 일반화 가능성 보장
- 민감도 분석(Sensitivity analysis): 연구 결과에 영향을 미치는 주요 변수들을 변동시켜가면서 결과의 변동을 관찰

$$Readmission_i = \lambda + \theta \cdot \widehat{SNFi} + \pi X_i + \eta_i$$

$$Readmission_i = \lambda + \theta \cdot SNFi + \pi_1 \cdot Physician_m + \pi_2 \cdot X_i + \eta_i$$

- ◆ 가정2. 도구변수(Physician_m) 가 오차항(η_i)과 상관관계가 없다는 것을 가정

- 도구와 오차항(η_i) 사이에 상관관계 0: 만약 도구 변수(Physician_m)의 계수가 유의미한 결과를 나타내지 않고 도구를 포함하여 추정결과가 크게 변하는 경우
- 도구와 오차항(η_i) 사이에 상관관계 X: 도구를 포함하더라도 추정 결과가 크게 변하지 않는 경우



36

가 통계적으로 유의미한 차이가 아니다

- ◆ Robustness Test(강건성 평가): 다양한 상황에서의 결과의 일관성과 안정성을 평가하는 것을 의미
→ 연구 결과의 신뢰성 및 일반화 가능성 보장
- 민감도 분석(Sensitivity analysis): 연구 결과에 영향을 미치는 주요 변수들을 변동시켜가면서 결과의 변동을 관찰

$$Readmission_i = \lambda + \theta \cdot \widehat{SNFi} + \pi X_i + \eta_i$$

$$Readmission_i = \lambda + \theta \cdot SNFi + \pi_1 \cdot Physician_m + \pi_2 \cdot X_i + \eta_i$$

	(1) 2SLS	(2) 2SLS with a readmission control
Panel A. Specification check		
Discharge to SNF	0.055*** (0.012)	0.036** (0.011)
Observations	132267	132267
Mean Y	0.029	
First stage F statistic	1244.143	1208.836

- ◆ 가정2. 도구변수(Physician_m) 가 오차항(η_i)과 상관관계가 없다는 것을 가정

- 도구와 오차항(η_i) 사이에 상관관계 0: 만약 도구 변수(Physician_m)의 계수가 유의미한 결과를 나타내지 않고 도구를 포함하여 추정결과가 크게 변하는 경우
- 도구와 오차항(η_i) 사이에 상관관계 X: 도구를 포함하더라도 추정 결과가 크게 변하지 않는 경우

- ◆ 결과 분석

- θ : 0.055 → 0.036
- 두 추정치의 차이가 통계적으로 유의미한 차이가 아니다

→ 가정2의 검증 완료

- 인과추론의 데이터과학. (2022, June 17). *[Bootcamp 3-3] 이/중차분법*.
<https://www.youtube.com/watch?v=yCeaZ9Ktk7g&t=404s>

Q&A

감사합니다.



KULSOM
오퍼레이션스 경영학회