



# 논문에서 사용하는 그림과 표의 올바른 이해

2023  
Jaelim Cho, MD, MPH, PhD

Assistant Professor  
Department of Preventive Medicine  
Yonsei University College of Medicine

# 연구 수행 과정

연구 단계	논문 섹션
연구 토픽에 대한 선행 연구 탐색을 통해 Research gap 확인 → 가설 설정	INTRODUCTION
가설에 걸맞는 연구 디자인 결정 변수 측정 방법 결정 연구 디자인에 맞는 통계 분석 방법 결정	METHODS Study design/Study population Measurements Statistical analysis
결과 보고	RESULTS
결과 해석	DISCUSSION

2023

Q. Research gap이란?

기존 연구에서 다루어지지 않은 부분,

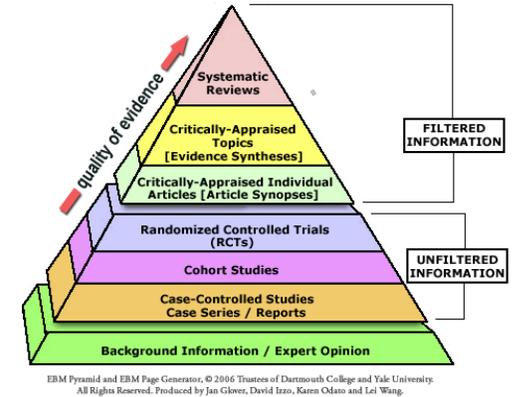
다뤄졌더라도 그 수가 매우 적거나 (little evidence), 결과가 일관되지 않아

(inconsistency; conflicting) 논란의 여지가 많은 (debatable; controversial) 부분,

기존 선행연구의 제한점 → 연구의 정당성, 필요성을 어필

# INTRODUCTION

- Literature review 시 systematic review (and meta-analysis) 먼저 찾기
  - 없으면, 최신 original article의 introduction 부분 참고
  - 없으면, 직접 literature review
- General to specific → previous studies → research gaps → rationale
- Research gaps
  - ✓ Little evidence, mixed/conflicting findings
  - ✓ Small sample size
  - ✓ Limited study populations (e.g., only children)
  - ✓ Less advanced study designs (e.g., cross-sectional study)
  - ✓ Less advanced techniques
  - ✓ Less advanced statistical methods



# METHODS

- Study design & study population
- Measurements (definitions of variables)
- Covariates (potential confounders)
- Statistical analysis – including sensitivity analysis

\*\* 일관성이 중요: literature review → research gaps → study aim → study design → statistical analysis → results → conclusions

\*\* 현재 연구를 통해 극복할 수 있는 부분이 research gap으로 강조되어야 함.

Mismatches.

- ✓ RCT가 부족한 실정 → 그런데 이 연구는 관찰연구?
- ✓ 기존 연구는 small sample size → 그러나 이 연구의 sample size도 비슷함?
- ✓ 목적에 제시되지 않은 가설이 갑자기 메인 통계 분석에 등장

# RESULTS

## Table 1. Characteristics

- Men accounted for 50% (n=xx).
- The proportion of men was 50% (n=xx)
- X was significantly higher in men than in women.
- Men were more likely to have a higher X compared with women.

## Table 2 or 3. Associations

- There was a positive/an inverse association between X and Y.
- X was positively/inversely associated with Y.
- X was significantly associated with a reduced/increased/higher Y.
- X was significantly associated with a reduced/increased/higher risk of Y.
- X increased the risk of Y.

\*\* Negative association = null association?

\*\* 도표의 모든 내용을 results에 작성할 필요는 없으나 유리한 결과만 골라서 쓰는 것은 바람직하지 않은 전략. 중립적인 관점에서 결과를 적는 것이 필요함 (No cherry-picking)

# DISCUSSION

- RESULTS는 틀에 박힌 방식으로 기술하고(copy & paste), DISCUSSION은 창의적인 방법으로 기술한다(paraphrasing).

(\*\*RESULTS에 paraphrasing이 많으면 독자가 헛갈릴 수 있다.)

- 구성

- Summary of findings: 연구 목적 간단히 + 결과 2-3개 정리
- 2-3개 bullet points → 한 주제 당 한 단락 (첫 문장이 topic sentence여야 함)
- Strengths: 독립 단락 또는 summary of findings에 끼워 넣어도 무방
- Limitations: 예상되는 문제나 바이어스 → 이를 해결하기 위해 나는 어떤 노력을 하였는가를 어필 → 해결 방법 없으면 향후 연구로...
- Conclusions: 연구 결과 요약 (+ implications, future studies)

# 가장 효과적인 도표의 선택

- 표: 구체적인 수치를 강조할 때
- 막대 그래프: 비교 수치(백분율)를 강조할 때
- 선 그래프: 추세를 강조할 때
- 내용이 허락하는 한 간결하게
- 색과 음영은 대조하고 싶을 때만 사용
- 아이콘 사용하지 말라(아마추어처럼 보일 수 있다)
- 표의 행과 열, 그래프의 양 축에 이름을 붙인다

출처: 케이트 트레이비언의 '시카고 양식' 제 8판 영어논문 바로쓰기 (시카고 대학 출판부 등 지음)

# Table 1. Descriptive statistics

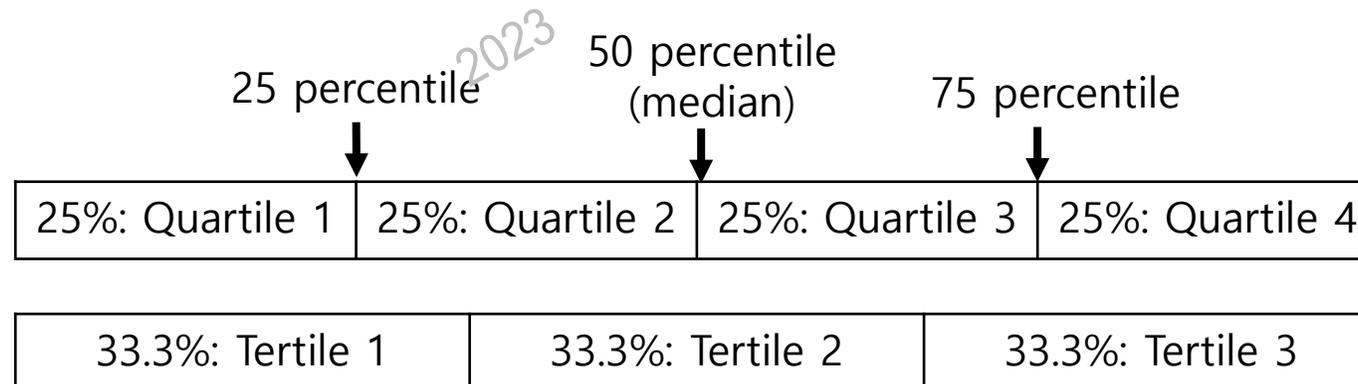
- 범주형 변수 (categorical variable)

: 빈도와 분율 (N, %)을 제시

- 연속형 변수 (continuous variable)

: 정규 분포인 변수는 Mean (standard deviation), 아닌 변수는 median (25%-75%)을 제시

Median 값은 분포에 따라 크게 변하지 않기 때문



# Table 1. Descriptive statistics

- Mean(SD)를 제시한 Table 1 변수에서 SD가 크면 정규분포하지 않을 가능성을 의심해야
- Example.

Triglyceride  $145 \pm 75$  mg/dL

95% within about 2SD above and below the mean

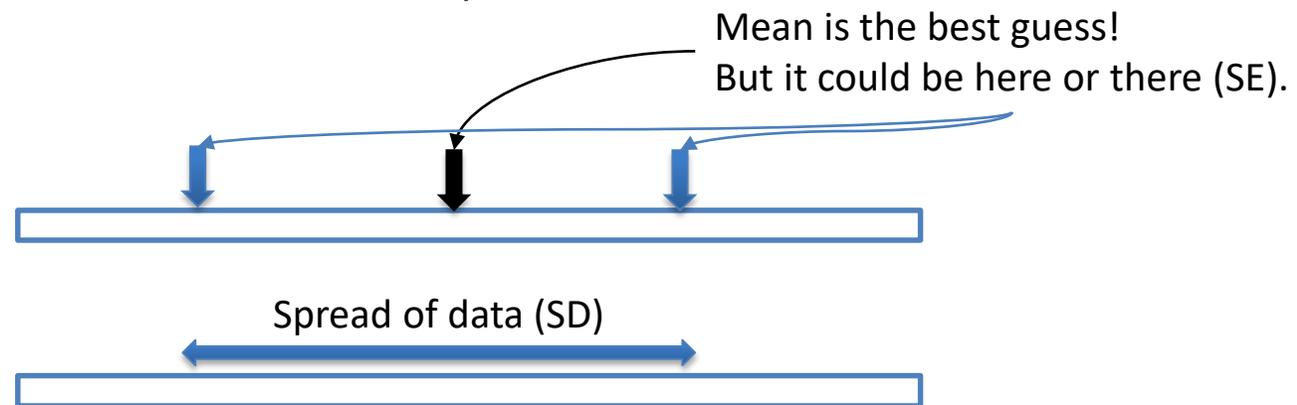
Upper limit:  $145 + (75 + 75) = 295$

Lower limit:  $145 - (75 + 75) = -5 \rightarrow \text{TG} = -5? \rightarrow \text{not normally distributed}$

2023

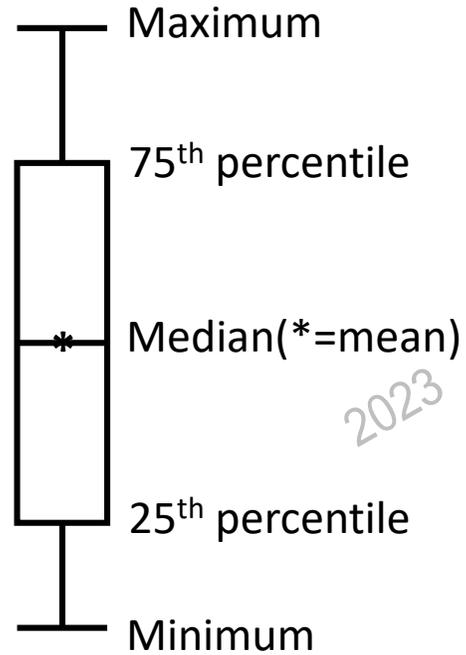
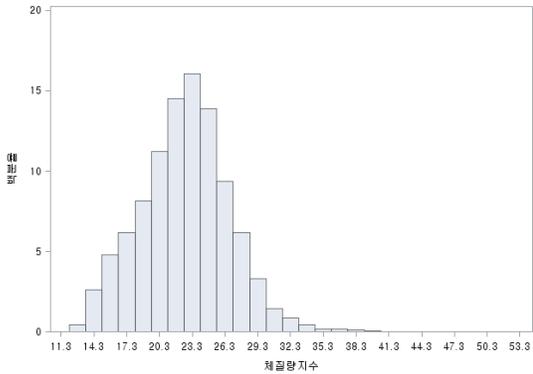
# SD and SE

- Standard deviation
    - 연속형 변수: mean and SD → 데이터 자체의 분포를 의미
  - Standard error
    - Sampling된 집단(e.g., KNHANES)에서 연속형 변수의 mean and SE
    - 선형회귀분석 결과: beta and SE (95% CI)
- 연관성 강도에 대한 가장 좋은 추정치는 beta이지만  $\pm 1.96 \times SE$  사이에 있을 수도 있다 (95% 자신함 = 5% 정도는 틀릴 수 있다고 인정)

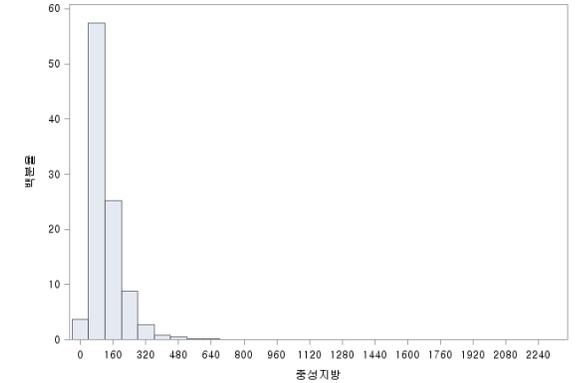
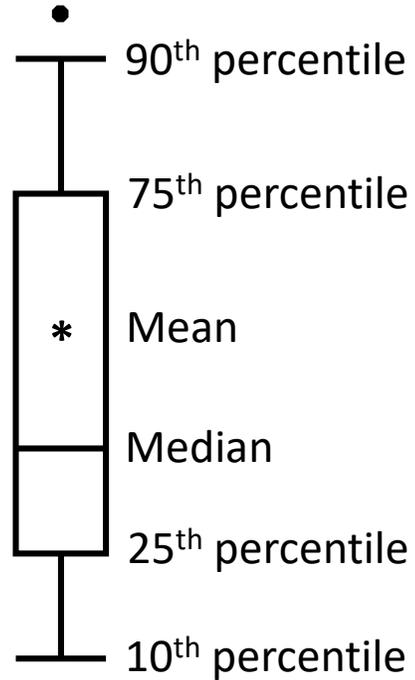




# 정규분포와 비정규분포



- Data points > 90<sup>th</sup> percentile



- 
- Data points < 10<sup>th</sup> percentile

Not skewed = normal distribution = parametric

Skewed = non-normal distribution = non-parametric

# 표 작성의 예 Table 1

**Table 1.** Characteristics of study participants.

Characteristics	Total (n=957)		Men (n=427)		Women (n=530)	
Age [mean (SD)]	67.3	(6.4)	68.3	(6.6)	66.4	(6.1)
Education level (y) [mean (SD)]	9.9	(4.3)	10.8	(4.3)	9.2	(4.2)
Marital status: cohabiting [n (%)]	804	(84.0)	403	(94.4)	401	(75.7)
Income per month <sup>a</sup> [n (%)]						
<\$455	95	(9.9)	36	(8.4)	59	(11.1)
\$455–\$908	123	(12.9)	53	(12.4)	70	(13.2)
\$909–\$1,363	134	(14)	52	(12.2)	82	(15.5)
\$1,364–\$1,817	151	(15.8)	69	(16.2)	82	(15.5)
\$1,818–\$2,726	164	(17.1)	74	(17.3)	90	(17.0)
\$2,727–\$3,635	110	(11.5)	62	(14.5)	48	(9.1)
\$3,636–\$5,454	89	(9.3)	44	(10.3)	45	(8.5)
≥\$5,455	46	(4.8)	24	(5.6)	22	(4.1)
Missing	45	(4.7)	13	(3.0)	32	(6.0)
History of disease [n (%)]						
Hypertension	313	(32.7)	149	(27.9)	194	(36.6)
Diabetes	170	(17.8)	89	(20.8)	81	(15.3)
Hyperlipidemia	308	(32.2)	101	(23.7)	207	(39.1)
Angina or myocardial infarction	89	(9.3)	42	(9.8)	47	(8.9)
Smoking [n (%)]						
Never smoker	633	(66.1)	119	(27.9)	514	(97.0)
Former smoker	261	(27.3)	249	(58.3)	12	(2.3)
Current smoker	63	(6.6)	59	(13.8)	4	(0.7)
Alcohol consumption [n (%)]	395	(41.3)	252	(59.0)	143	(27.0)
Walking days per week [mean (SD)]	4.7	(2.3)	4.7	(2.3)	4.7	(2.3)
Body mass index (kg/m <sup>2</sup> ) [mean (SD)]	24.7	(3)	24.7	(2.8)	24.7	(3.1)
Systolic blood pressure (mmHg) [mean (SD)]	128.5	(14.2)	129.2	(13.3)	128.0	(14.9)
Diastolic blood pressure (mmHg) [mean (SD)]	75.2	(9.1)	75.9	(8.6)	74.7	(9.4)
Total cholesterol (mg/dL) [mean (SD)]	183.2	(37.1)	175.9	(38.4)	189.1	(35.0)
Fasting blood glucose (mg/dL) [mean (SD)]	99.9	(21.5)	102.5	(21.5)	97.8	(21.2)

- 기술 통계량
- ✓ 정규분포 연속 변수는 mean (SD)
- ✓ 비정규분포 연속 변수는 median (25%-75%)
- ✓ 빈도 변수는 N(%)

# 표 작성의 예 Table 1

Table 1. Characteristics of study participants at baseline

Variables	Total (N=337)	Never smokers (N=215)	Ever smokers (N=122)	P
<b>Age, mean (SD)</b>	67.5 (6.1)	67.3 (5.7)	67.9 (6.7)	0.38
<b>Sex, N (%)</b>				<0.001
Male	164 (48.7)	46 (21.4)	118 (96.7)	
Female	173 (51.3)	169 (78.6)	4 (3.3)	
<b>Years of education, mean (SD)</b>	10.6 (4.2)	10.09 (4.1)	11.37 (4.2)	0.007
<b>Body mass index, mean (SD)</b>	24.4 (2.8)	24.34 (2.8)	24.51 (2.8)	0.60
<b>Smoking status, N (%)</b>				<0.001
Never smoker	215 (63.8)	215 (100.0)	0 (0.00)	
Former smoker	105 (31.2)	0 (0.00)	105 (86.1)	
Current smoker	17 (5.0)	0 (0.00)	17 (13.9)	
<b>Alcohol consumption, N (%)</b>				<0.001
No	202 (59.9)	150 (69.8)	52 (42.6)	
Yes	135 (40.1)	65 (30.2)	70 (57.4)	
<b>Blood pressure status, N (%)</b>				0.068
Normal	156 (46.3)	96 (44.7)	60 (49.2)	
Undiagnosed hypertension	24 (7.1)	11 (5.1)	13 (10.7)	
Diagnosed hypertension	157 (46.6)	108 (50.2)	49 (40.2)	
<b>Glycemic status, N (%)</b>				0.075
Normoglycemia	178 (52.8)	125 (58.2)	53 (43.4)	
Prediabetes	70 (20.8)	40 (18.6)	30 (24.6)	
Undiagnosed diabetes mellitus	19 (5.6)	10 (4.7)	9 (7.4)	
Diagnosed diabetes mellitus	70 (20.8)	40 (18.6)	30 (24.6)	
<b>Urinary PAH metabolites (µg/g creatinine), median (25%–75%)</b>				
1-OHP	0.19 (0.14–0.33)	0.17 (0.12–0.30)	0.21 (0.16–0.34)	0.007
2-Naphthol	1.78 (0.41–4.49)	1.74 (0.45–4.36)	2.03 (0.26–4.85)	0.98
1-OHPHE	0.09 (0.05–0.16)	0.08 (0.05–0.16)	0.09 (0.06–0.17)	0.59
2-OHF	0.16 (0.10–0.25)	0.15 (0.09–0.23)	0.18 (0.11–0.43)	0.005
<b>Intracranial volume (1,000mm<sup>3</sup>), mean (SD)</b>	1,522.6 (165.2)	1,471.5 (145.2)	1,612.6 (160.3)	<0.001

Abbreviations. SD = standard deviation; PAH = polycyclic aromatic hydrocarbon; 1-OHP = 1-hydroxypyrene; 1-OHPHE = 1-hydroxyphenanthrene; 2-OHF = 2-hydroxyfluorene. Footnote. Differences in characteristics between ever and never smokers were tested by two-sample t-test, chi-square test, or Wilcoxon Rank Sum test.

- 기술 통계량
- ✓ 정규분포 연속 변수는 mean (SD)
- ✓ 비정규분포 연속 변수는 median (25%-75%)
- ✓ 빈도 변수는 N(%)
- P-value (두 그룹 차이의 유의성)
- ✓ 정규분포 연속 변수는 t-test
- ✓ 비정규분포 연속 변수는 Mann-Whitney U test
- ✓ 빈도 변수는 chi-squared test

# 표 작성의 예 Association 결과 reporting

**Table 2.** Association between air pollution and brain magnetic resonance imaging markers in adults.

Pollutant	Measurement	Brain region	Model 1			Model 2		
			Beta	(95% CI)	p-Value	Beta	(95% CI)	p-Value
PM10	Cortical thickness (mm)	Global	0.00	(-0.01, 0.01)	0.60	0.00	(-0.01, 0.01)	0.60
		Frontal	-0.02	(-0.03, -0.01)	0.003	-0.02	(-0.03, -0.01)	0.005
		Parietal	0.00	(-0.02, 0.01)	0.68	-0.01	(-0.02, 0.01)	0.45
		Temporal	-0.06	(-0.07, -0.05)	<0.001	-0.06	(-0.07, -0.04)	<0.001
		Occipital	0.05	(0.04, 0.07)	<0.001	0.05	(0.03, 0.06)	<0.001
		Cingulate	0.04	(0.02, 0.05)	<0.001	0.03	(0.02, 0.05)	<0.001
		Insula	-0.02	(-0.04, 0.00)	0.015	-0.02	(-0.04, 0.00)	0.075
	Subcortical volume (mm <sup>3</sup> )	Thalamus	-192.7	(-263.1, -122.3)	<0.001	-184.0	(-255.4, -112.5)	<0.001
		Caudate	-34.4	(-91.2, 22.5)	0.24	-45.4	(-103.6, 12.8)	0.13
		Putamen	-46.2	(-107.3, 14.8)	0.14	-60.6	(-122.4, 1.3)	0.055
		Pallidum	-49.8	(-76.3, -23.2)	<0.001	-49.3	(-76.4, -22.1)	<0.001
		Hippocampus	-58.4	(-104.2, -12.7)	0.012	-55.4	(-101.6, -9.3)	0.019
		Amygdala	-38.9	(-62.0, -15.8)	0.001	-36.2	(-59.7, -12.6)	0.003
		Nucleus accumbens	-63.1	(-76.2, -50.0)	<0.001	-56.8	(-70.1, -43.5)	<0.001

- Linear regression: beta (SE or 95% CI) + p-value      Beta 95% CI crossing 0 = null
- Logistic regression: odds ratio (95% CI)      OR 95% CI crossing 1 = null
- Cox regression: hazard ratio (95% CI)      HR 95% CI crossing 1 = null

# 표 작성의 예 Association 결과 reporting

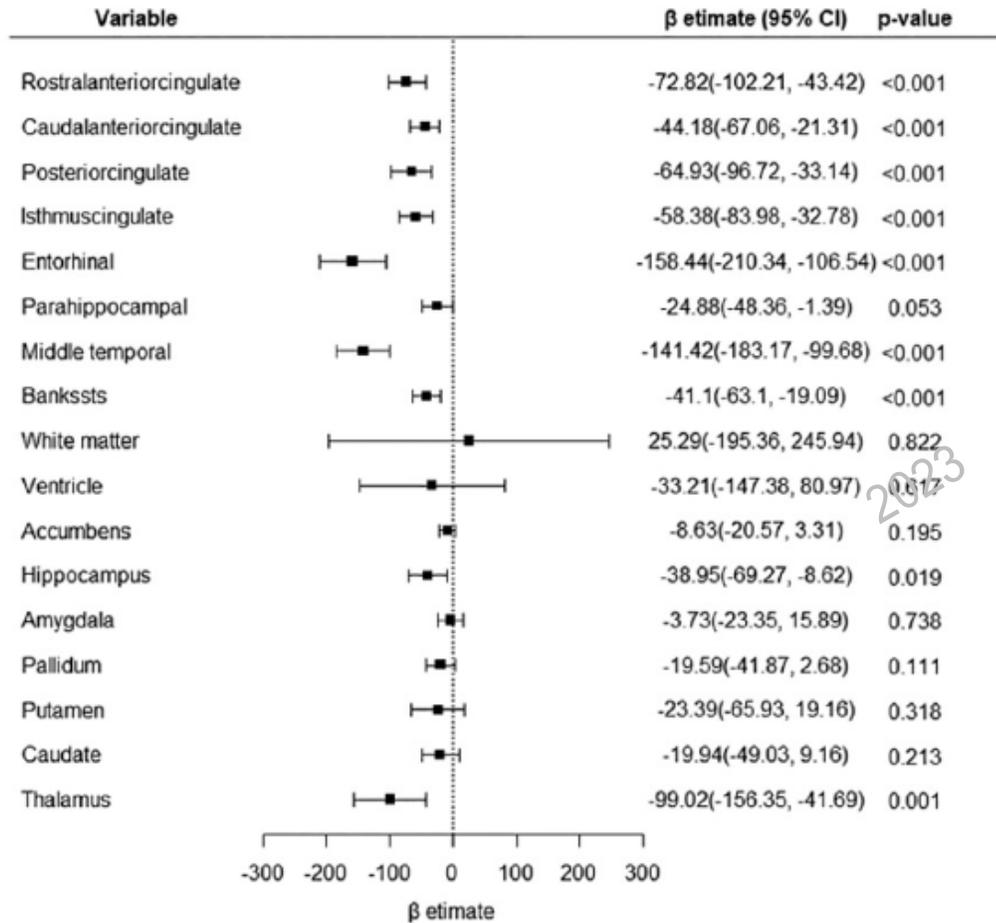
Table 2—Associations between the use of antidiabetic medications and mortality in individuals with DEP and type 2 diabetes

	Mortality rate per 100 person-years	Crude HR (95% CI)	Adjusted HR (95% CI)
Insulin use			
PPDM			
No antidiabetic medications	16.47 (12.78–20.15)	1.00	1.00
Never insulin	5.87 (4.42–7.31)	0.37 (0.27–0.51)	0.51 (0.36–0.71)
Ever insulin	5.20 (3.51–6.89)	0.34 (0.23–0.50)	0.75 (0.49–1.14)
PCRD			
No antidiabetic medications	126.91 (95.64–158.19)	1.00	1.00
Never insulin	65.49 (56.74–74.24)	0.52 (0.44–0.62)	0.59 (0.49–0.71)
Ever insulin	49.06 (42.99–55.12)	0.41 (0.35–0.49)	0.46 (0.39–0.55)
Type 2 diabetes			
No antidiabetic medications	9.67 (9.42–9.92)	1.00	1.00
Never insulin	5.05 (4.96–5.15)	0.53 (0.52–0.55)	0.80 (0.78–0.83)
Ever insulin	4.10 (3.98–4.21)	0.45 (0.43–0.47)	0.86 (0.83–0.90)

- Linear regression: beta (SE or 95% CI) + p-value
- Logistic regression: odds ratio (95% CI)
- Cox regression: hazard ratio (95% CI)

Beta 95% CI crossing 0 = null  
 OR 95% CI crossing 1 = null  
 HR 95% CI crossing 1 = null

# 그림 작성의 예 Association 결과 reporting



Forest plot from Kim YT et al., STOTEN, 2022

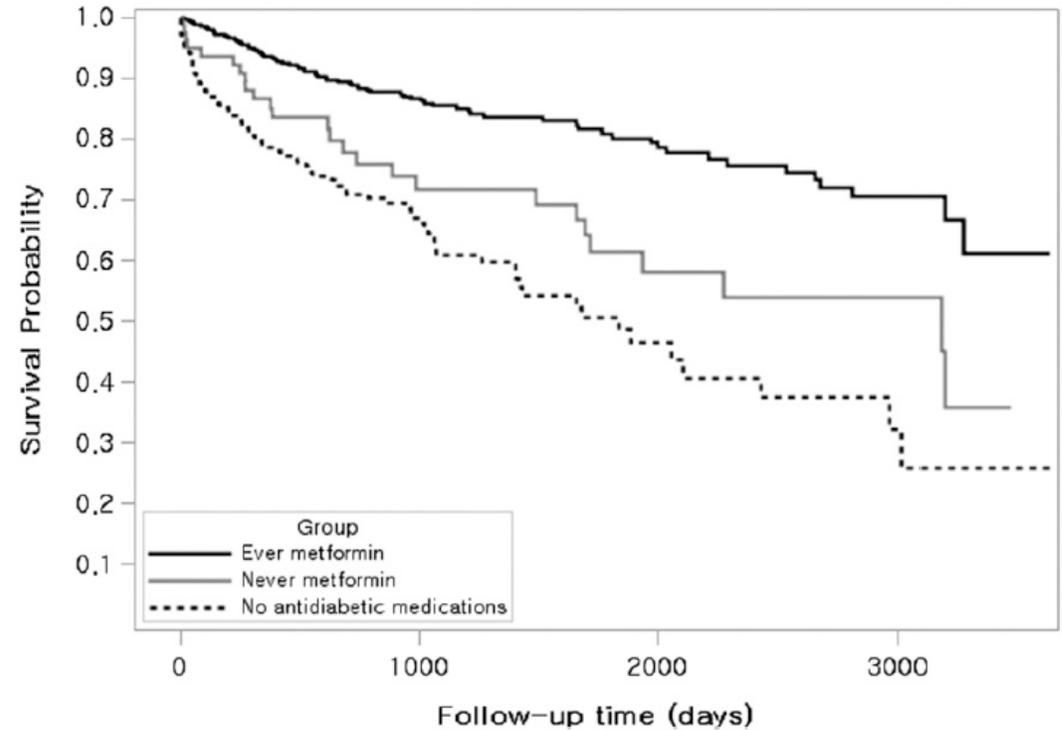


Figure 1—Survival probability of individuals with PPDM stratified by the use of antidiabetic medications.

Kaplan-Meier curve from Cho J et al., Diabetes Care, 2019

# 도표 이름 짓기

- 포괄적인 제목이나 설명은 피한다.
  - 명사구 사용. 관계사절을 피하고 대신 분사구를 사용  
(Number of families that subscribe... (x), Number of families subscribing... (o))
  - 배경지식이나 자료 의미를 설명하지 않는다.
  - 비슷한 자료를 제시하는 도표와 구분 가능한 이름 붙이기
- \*\* 연구논문에서의 도표는 본문과 독립적으로 (stand-alone) 구체적인 정보를 전달해야 한다.

## Examples.

- ① OR of dementia and medication
- ② Odds ratios of dementia
- ③ Associations between medication and dementia
- ④ Associations between antidiabetic medications and the risk of Alzheimer's disease in patients with type 2 diabetes
- ⑤ Risks of Alzheimer's disease associated with the use of antidiabetic medications in patients with type 2 diabetes

2023

# Formatting checklist

- 타이틀: 표는 위, 그림은 아래
- 표 가로 선은 맨 위 두 개와 맨 아래만 (중간 가로선은 없게끔)
- 동일한 단어 반복을 지양하도록 세부 컬럼 구성
- 숫자와 숫자 사이는 긴 하이픈(-), 그 외는 짧은 하이픈(-) (마이너스, 문자/숫자와 문자 사이)
- P value는 저널마다 상이하나 보통 최대 소숫점 셋째 자리까지만 표시, 0.001보다 작으면  $p < 0.001$ 로 표기
- Table 1 백분율은 컬럼 백분율 사용<sup>2028</sup> 집단의 분포를 보여주기 위한 것이므로
- 셀 공간이 충분하면 약자 사용을 지양함
- 약자와 통계 방법, 모델 공변량은 표 아래에 footnote로 반드시 작성
- 통계적으로 유의한 것은 bold 처리 해도 되고 안 해도 됨(대개 안 함)



도표는 연구 결과를 효율적이고 효과적으로 전달할 수 있는 수단  
도표에 대한 이해는 독자에 대한 이해가 전제되어야

**THANK YOU FOR YOUR ATTENTION!**