

자료의 분석과 표현



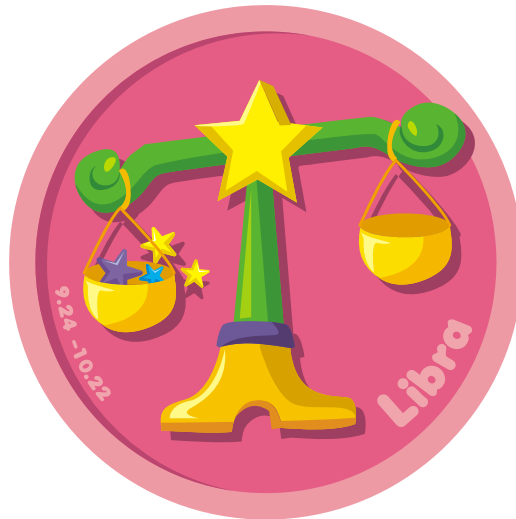
2009. 3. 27.

서울의대 예방의학교실 박병주



비과학적 연구와 연구윤리

**“비과학적으로 수행된 연구는
비윤리적이다.”**



과학적이고 타당한 연구결론을 얻으려면? : 연구계획

연구가설

- 연구가설을 구체적으로 수립

연구설계

- 수립된 가설을 검증할 수 있는 적절한 연구설계 선택

연구계획서 작성

- 연구수행과정을 구체적, 객관적으로 기술한 연구계획서 개발

과학적이고 타당한 연구결론을 얻으려면?

: 연구수행

연구대상 확보


- 연구목적에 적합한 대상 확보

자료수집

- 정확한 연구자료를 누락없이 수집

데이터베이스 구축

- 연구자료를 오류 없는 완전한 **DB**로 구축



과학적이고 타당한 연구결론을 얻으려면? : 연구분석 및 결과제시

자료분석

- 적절한 통계분석법 적용

결과해석 및 결론

- 정확한 해석, 타당한 결론

논문 작성

- 적절한 표와 그림으로
결과제시



임상연구와 비뚤림(Bias)

❖ 계통적 오류 (Systematic Error, Bias);

연구수행상의 특별한 이유에 의한 비뚤림

1. 선택비뚤림
2. 정보비뚤림
3. 교란비뚤림



선택비뚤림(Selection bias)

- ❖ 연구대상이 모집단을 **대표**하지 못하거나, 연구 목적에 적합하지 **않는** 대상을 선정함으로써 타당한 결론을 내리지 못하는 경우

새로운 **ACE Inhibitor**계 항고혈압제의 효과 평가 시

- 연구대상자 가운데 만성 신 질환자가 다수 포함되는 경우



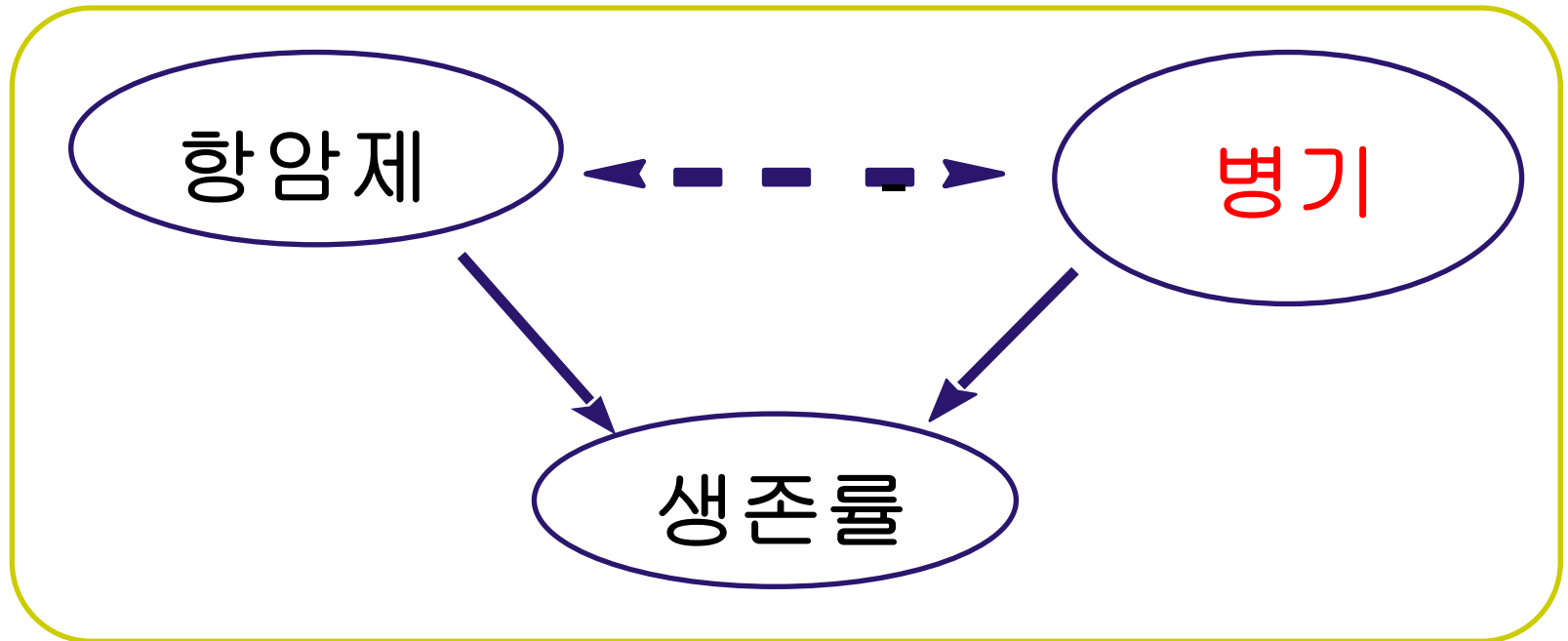
정보비뚤림(Information bias)

❖ 연구자료를 수집하는 과정에 발생할 수 있는 비뚤림

새로운 진통제의 효과를 위약과 비교하여 평가할 때
- 연구대상자들이 자신에게 투여되는 약이 신약인지 위약인지 알고 있어 효과에 영향을 미칠 때

교란비동림(Confounding)

새로운 항암제와 표준 항암제의 생존률 비교시





과학적으로 타당한 연구

의학연구를 과학적으로 수행하는 방법을
제대로 알지 못한 상태에서
연구를 수행, 잘못된 결론을 내리고 발표



다른 연구자는 물론 국민들에게도 피해!!



의학연구에서 통계학의 역할

❖ 적용범위

- 연구계획: 연구설계, 연구규모 추정
- 연구수행 후 통계분석
- 논문작성

오용되고 잘못 해석될 위험성!

통계적 개념 정확히 이해
고급통계방법 적용시 통계전문가들과 협력



통계분석의 기본개념

❖ 변수

- 개인에 따라 다르게 관찰되는 특성.
- 크게 양적 변수와 질적 변수로 구분.

❖ 확률분포

- 동일 집단 또는 현상의 반복측정 결과가 나타내는 집합적 양상

❖ 추정과 검정

- 본집단의 한정된 자료를 이용하여 일반 모집단의 현상을 추정
- 검정을 통해 일반화

❖ 양적 변수

- 특정한 숫자로 관찰되는 변수
- 신장, 체중, 심박수 등
- 이산형 변수 혹은 연속형 변수로 구분

❖ 질적 변수

- 특정한 범주로 관찰되는 변수
- 성별, 종교 등

- ❖ 동일 집단 또는 현상의 반복측정 결과가 나타내는 집합적 양상
- ❖ 관측결과에 따른 분포
 - 명목형 변수인 경우 이산분포
 - 연속형 변수인 경우 연속분포

중심극한정리

모집단이 정규분포를 따르지 않더라도 모집단에서 무작위로 반복하여 추출된 표본의 평균은, 표본의 크기가 어느 정도 크면 정규분포를 따름.

❖ 추정(Estimation)

- 점 추정치(Point Estimation)
 - 모수를 추정하기 위해 쓰이는 수치
- 구간추정치(Interval Estimation)
 - 일정 수준의 신뢰도로 모수를 포함하는 구간

❖ 검정의 목적(Hypothesis Testing)

- 표본에서 관찰된 현상을 바탕으로 모집단에서 발생하는 현상에 대한 결론을 내리고자 하는 것
 - 귀무가설(H_0)
 - 대립가설(H_1)

제1종 오류와 제2종 오류

■ 가설의 검정

		실제로 귀무가설	
		참	거짓
검정결과 귀무가설	채택	옳음	제 2종 오류 ($1-\beta$)
	기각	제 1종 오류 (α)	옳음; 검정 력 (β)

■ α, β 의 결정기준은?

- ➔ 절대적 기준은 없다!
- ➔ 연구자가 주관적으로 결정

일반적으로
 $\alpha = 0.01$ or 0.05
 $\beta = 0.1$ or 0.2



통계적 검정력 유의수준과 p-value

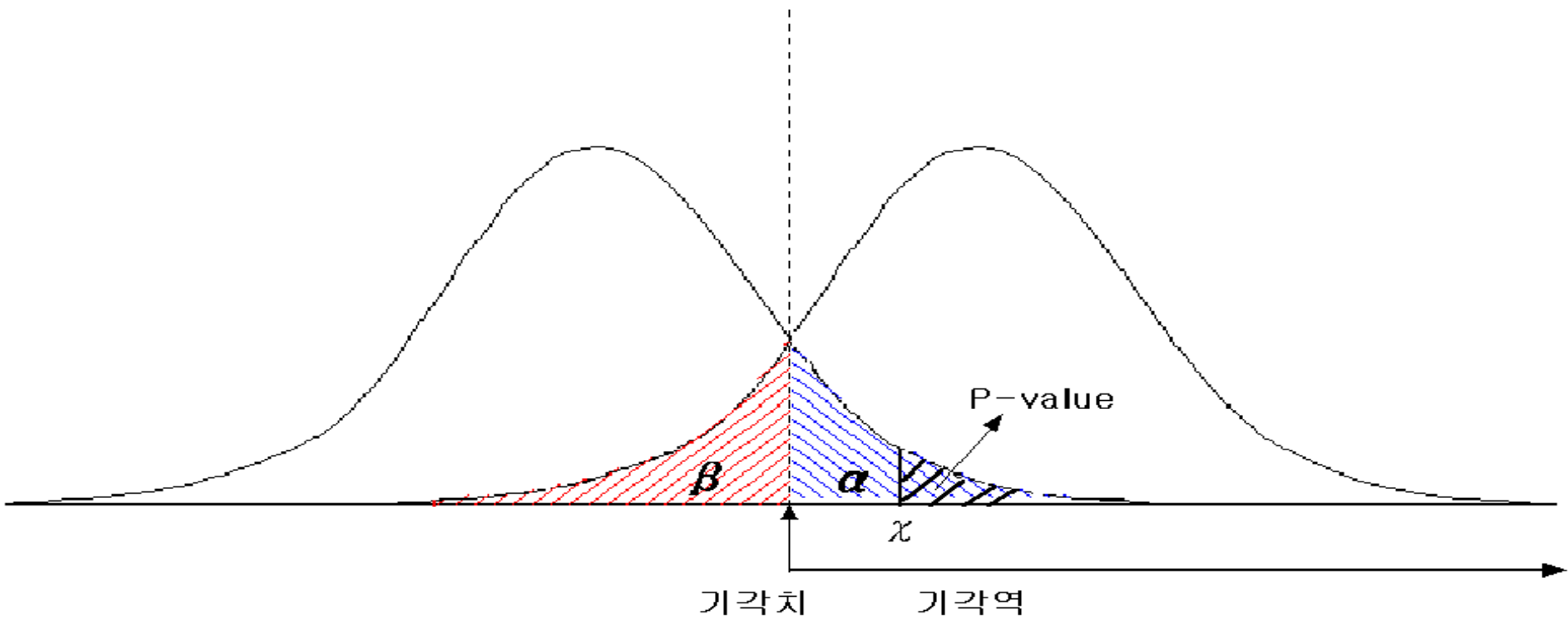
❖ 통계적 검정력(Statistical Power, $1-\beta$)

- 연구조사 대상 집단의 규모가 부적절하게 작은 경우, 통계적 검정력이 떨어져 위음성으로 판정할 오류 증가

❖ 유의수준(significance level)과 p값(p-value)

- 유의수준: 제1종오류를 범할 확률의 최대허용한계
- p값: 귀무가설이 참일 때, 표본에서 계산된 검정통계량보다 클(혹은 작을) 값이 나올 확률
- **p-value ≈ 0.0** 은 귀무가설이 참일 때, 임의로 표본을 추출하였을 때 그러한 표본 평균이 나올 확률이 거의 0이라는 의미.

통계 용어의 정리





신뢰구간(Confidence Interval)

▶ 신뢰구간

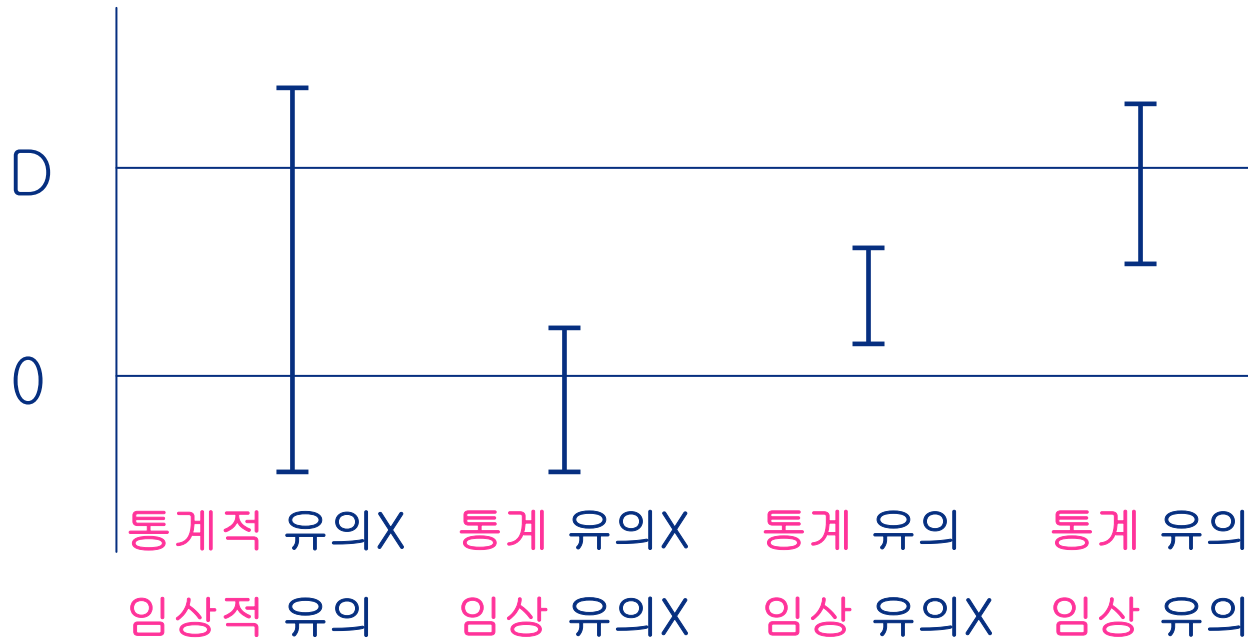
- ❖ 표본을 대상으로 측정한 값으로부터 모집단의 대표값(모수)이 위치할 구간을 확률적으로 추정한 범위

Ex . 95% 신뢰구간 추정의 의미는 ?

→ 신뢰구간 내에 모수가 포함될 확률이 **95%**이다!

▶ *P-값보다 신뢰구간을 제시하는 것이 더 좋다!*

통계적 유의성 임상적 유의성



임상적으로
유의한
최소크기

치료효과크기 < D → 치료법 동등

통계적 유의성 ≠ 임상적 유의성



통계분석의 원칙

❖ 통계분석의 원칙, 구체적인 분석방법

→ 연구계획서에 미리 기술

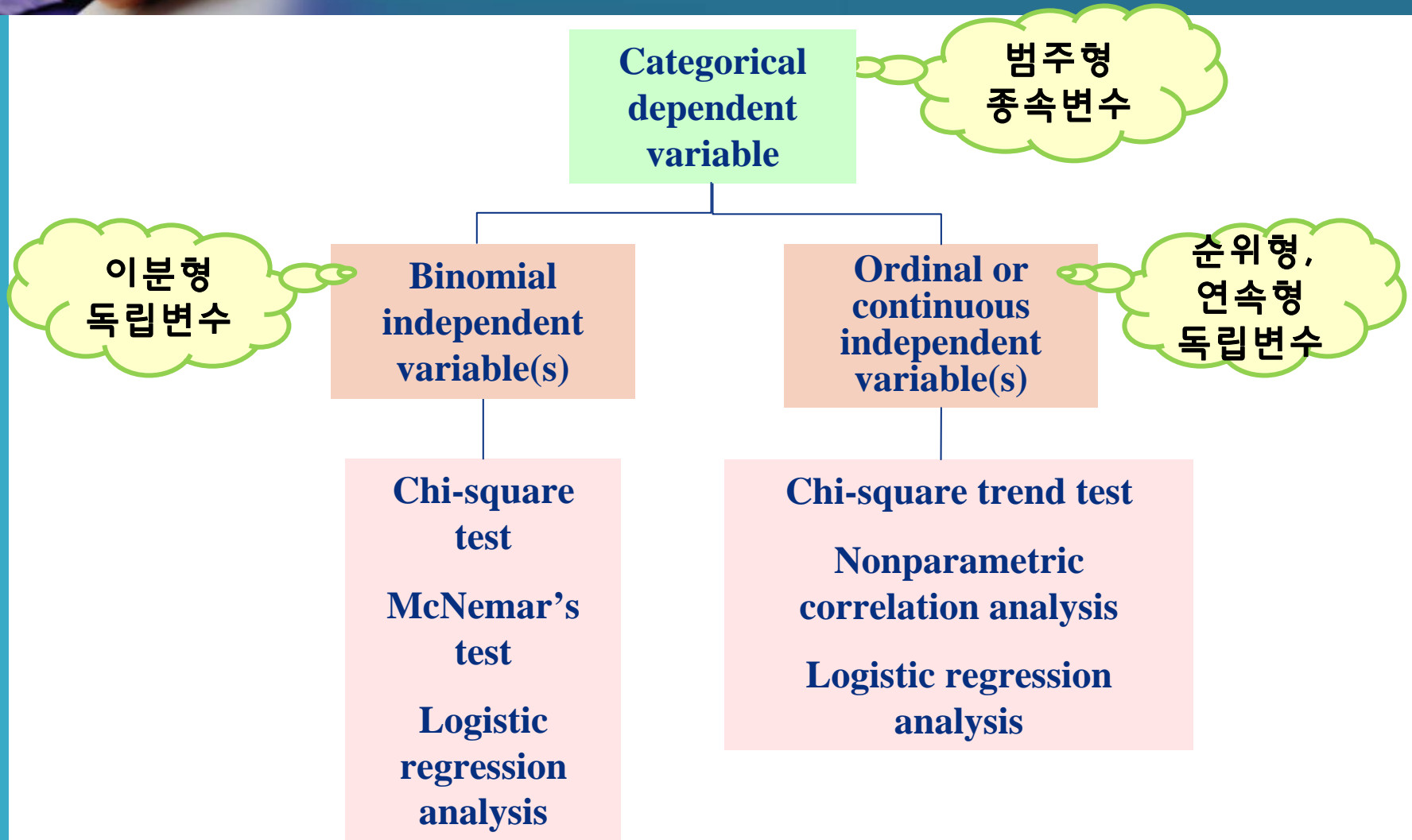
- 각 평가변수에 대한 분석방법
- 분석대상군의 정의
- 예상되는 분석상의 문제점을 해결하는 방법 등



통계분석방법의 선택

- ❖ 각 상황에 따라 가장 적절한 통계분석방법 결정!
- ❖ 검정하고자 하는 **변수의 척도**: 평균값, 빈도
- ❖ 검정하고자 하는 **가설**: 차이성, 우위성, 동등성, 비열등성
- ❖ **두 개 이상의 집단을** 대상으로 하는 경우: **독립성 여부**
- ❖ **단변량분석, 다변량분석** 시행 여부

통계분석방법의 선택



통계분석방법의 선택

연속형
종속변수

Continuous
dependent
variable

명목형
독립변수

Nominal
independent
variable(s)

t-test
(Mann-Whitney)
paired t-test
(Wilcoxon)
ANOVA
(Kruskal-Wallis)

Continuous
independent
variable(s)

Correlation &
Regression
analysis

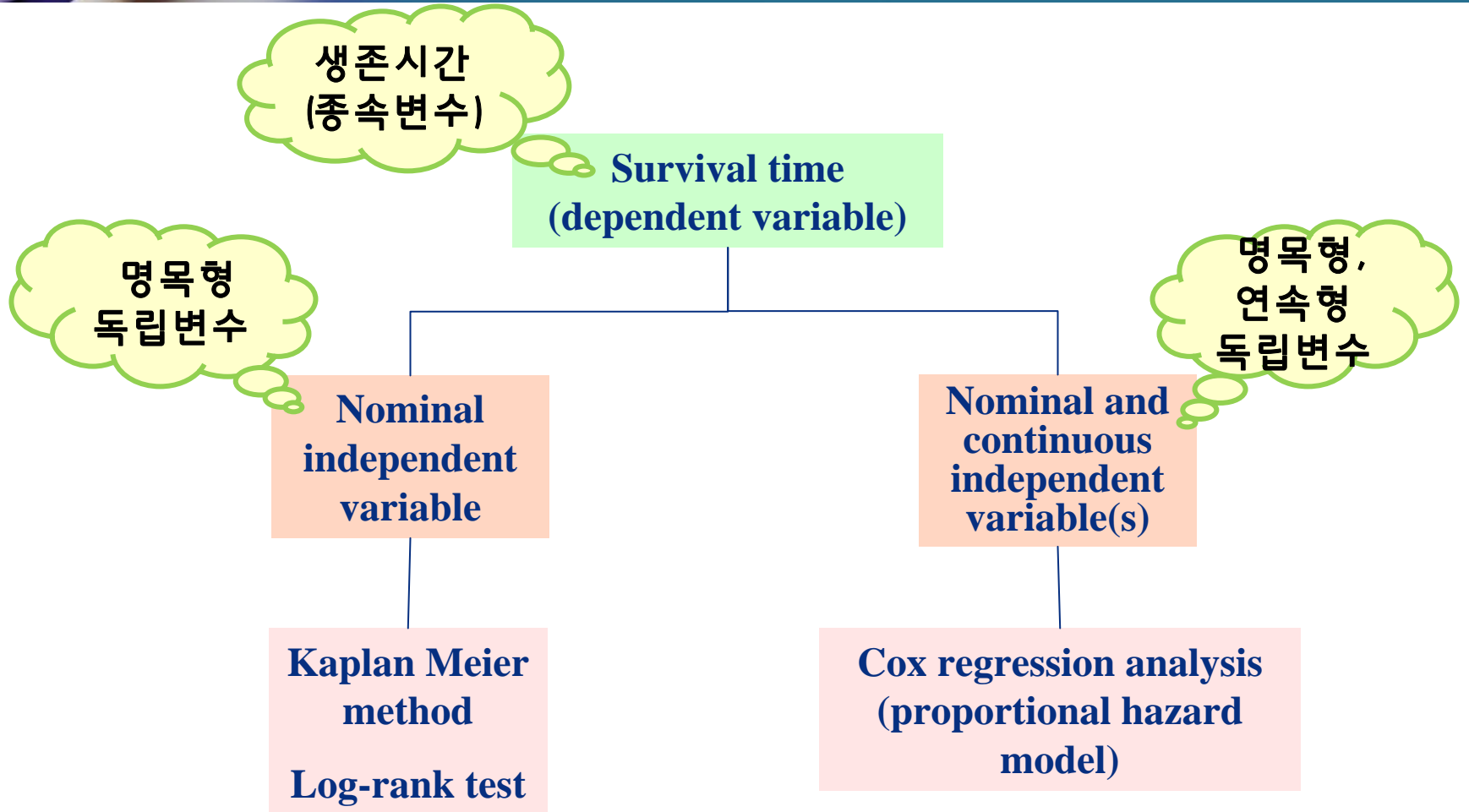
연속형
독립변수

Nominal and
continuous
independent
variable(s)

ANCOVA
Repeated
Measures
ANOVA
Regression
analysis

명목형,
연속형
독립변수

통계분석방법의 선택



❖ 분석군의 정의

- **ITT (Intention-to-Treat):** 의도(배정)된대로 분석
 - 무작위배정된 모든 피험자들을 통계분석에 포함

■ **PP (Per-Protocol):** 계획서순응군분석

- 임상시험계획서를 준수한 경우
- 실제 치료받은 내용에 따라 분석

분석대상군

Intent-to-treat (ITT) 환자군은 실제로 시험약을 투여받았는지에 관계 없이 시험약에 무작위배정된 모든 환자로 구성된다. 이 환자군은 환자가 무작위배정된 투여군을 바탕으로 하며 유효성 자료 분석시 보조 환자군으로 사용된다.

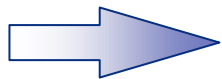
수정된 Intent-to-treat 환자군(MITT)은 ITT 환자군 중 (FISH)증양 검사에 의해 ER 양성으로 확인된 모든 환자로 구성된다. 이는 유효성 자료 분석시 주 환자군이 된다.

Per-protocol (PP) 환자군은 무작위배정되어 약제를 복용한 환자 중 임상시험계획서에 순응한 모든 환자로 구성된다. 환자를 PP 환자군에서 제외하는 중대한 임상시험계획서 위반에 대해서는 맹검 해제 전에 RAP에 정의한다. PP 환자군은 질병진행까지의 시간에 대한 보충 분석에 사용된다.

안전성 분석 환자군은 임상적 안전성과 내약성을 평가하는데 사용되며 무작위배정되어 시험약을 1 회 이상 복용한 모든 환자로 구성된다.

❖ 통계분석은 연구가설의 유형에 따라 달라짐

- 차이성 검정 : 두 군의 치료효과가 차이가 있는지 여부를 검정
- 우위성 검정 : 새로운 치료제가 기존의 치료제보다 더 우수하다는 것을 검정
- 비열등성 검정 : 새로운 치료제의 치료효과가 기존의 표준치료제에 비해 나쁘지 않다는 것 검정
- 동등성 검정 : 두 군에서 치료효과의 차이가 임상적으로 무시할 수 있는 동등성 인정한계 내에 있는지를 검정



연구목적에 부합하는 연구가설을 세우고 분석을 시행하는 것이 중요함.



결과: 표와 그림 제시

❖ **Self-explanatory!**

❖ 제목

- What, Who, Where, When

❖ 각주

- 약어(**Abbreviation**) 또는 기호를 설명
- 출처를 밝힐 것!



Table 1. General characteristics

Table 1. Characteristics of the study patients and prescriptions 2001, Busan

		Patients		Prescriptions	
		No.	(%)	No.	(%)
Age	65-69	110,961	(42.2)	1,044,068	(42.0)
	70-74	76,266	(29.0)	761,028	(30.6)
	75-79	45,661	(17.4)	437,346	(17.6)
	80-84	20,906	(8.0)	177,667	(7.2)
	85+	9,162	(3.5)	63,118	(2.5)
Gender	Male	92,951	(35.3)	829,175	(33.4)
	Female	170,005	(64.7)	1,654,052	(66.6)
Total		262,956	(100.0)	2,483,227	(100.0)

발생률 및 사망률 : 점추정치와 신뢰구간

Table 2. Age-specific incidence rates of lung cancer per 100,000 person-years in male in Korean Elderly Pharmacoepidemiologic Cohort (KEPEC), Busan, 1994-1998

Age	Observed period(person-years)	No. of incident cases	Incidence rate per 100,000 Person-years (95% CI)
65-69	24,552	73	297.3 (229.1 - 365.5)
70-74	13,503	53	392.5 (286.8 - 498.2)
75-79	7,189	19	264.3 (145.5 - 383.1)
80-84	3,029	10	330.1 (125.5 - 534.8)
85+	1,112	1	89.9 (0.0 - 266.2)
Total	49,385	156	315.9 (266.3-365.5)

Age adjusted rate for the Korean population as of 1995(95% CI) = 316.9 (267.2-366.6)

Age adjusted rate for the world population(95% CI) = 307.3 (259.1-355.5)

Table 6. Age-specific mortality rates of lung cancer by sex per 100,000 person-years in Korean Elderly Pharmacoepidemiologic Cohort (KEPEC), Busan, 1994-1998

Age	Male	Female
	Mortality rate per 100,000 pys (95% CI)	Mortality rate per 100,000 pys (95% CI)
65-69	146.6 (98.7-194.5)	44.7 (23.4 - 65.9)
70-74	488.8 (370.9-606.7)	86.4 (47.6 - 125.3)
75-79	500.8 (337.2-664.3)	137.9 (75.9 - 199.9)
80-84	495.2 (244.6-745.8)	118.0 (44.9 - 191.2)
85+	449.6 (55.5-843.8)	111.0 (13.7 - 208.3)
Crude rate	319.8 (270.0-369.8)	80.7 (61.8 - 99.6)

Age adjusted rate for the Korean population as of 1995 in male(95% CI)= 342.3 (288.9-395.7)

Age adjusted rate for the world population in male(95% CI)= 341.5 (288.3-394.7)

Age adjusted rate for the Korean population as of 1995 in female(95% CI)= 84.8 (64.9-104.7)

Age adjusted rate for the world population in female(95% CI)= 79.9 (61.2-98.6)

환자-대조군연구 : OR and 95% CI

Table 2 Association between PPA and risk of HS in all subjects

Variable	Cases (n = 940), n (%)	Controls (n = 1,880), n (%)	Unadjusted OR (95% CI)	Adjusted OR* (95% CI)
Exposure to PPA†				
No	924 (98.3)	1,866 (99.3)	1.00	1.00
Yes	16 (1.7)	14 (0.7)	2.46 (1.15–5.24)	2.14 (0.94–4.84)
Window of exposure to PPA‡				
No	924 (98.3)	1,866 (99.3)	1.00	1.00
3 Days or less	10 (1.1)	4 (0.2)	6.27 (1.71–22.98)	5.36 (1.40–20.46)
4–14 Days	6 (0.6)	10 (0.5)	1.28 (0.46–3.55)	1.04 (0.33–3.21)
p Value¶			0.004	0.02
Duration of exposure to PPA§				
No	924 (98.3)	1,866 (99.3)	1.00	1.00
1–2 Days	7 (0.7)	9 (0.5)	1.61 (0.58–4.49)	1.38 (0.44–4.32)
3 Days or more	9 (1.0)	5 (0.3)	4.16 (1.27–13.64)	3.44 (1.01–11.67)
p Value¶			0.009	0.04
Dose of PPA				
No	924 (98.3)	1,866 (99.3)	1.00	1.00
<75 mg per day	7 (0.7)	6 (0.3)	2.43 (0.82–7.27)	2.13 (0.64–7.12)
≥75 mg per day	9 (1.0)	8 (0.4)	2.48 (0.91–6.74)	2.14 (0.74–6.14)
p Value¶			0.03	0.08

* Odds ratio (OR) and 95% CI are calculated by conditional logistic regression and adjusted for age group (defined by decades), a history of hypertension, a family history of stroke, upper respiratory tract infections within 30 days before index date, and alcohol consumption.

† Exposure to PPA was defined as being exposed to PPA within 14 days before the index date.

‡ Window of exposure to PPA was defined as the interval from the last exposure date to the index date.

¶ p Value is calculated by the likelihood ratio test for trend to investigate the dose-response relationship between variables and risk of hemorrhagic stroke.

§ Duration of exposure to PPA was defined as days of taking PPA-containing drugs within 14 days before the index date.

PPA = phenylpropanolamine; HS = hemorrhagic stroke.

코호트연구: RR and 95% CI

Table 4. Rate Ratios for Mortality by Time from Last Combined Prescription to Time of Death Among New Users of Cisapride

Variable	Person-years	Subjects, n	Deaths, n	Crude RR (95% CI)	Adjusted RR ^a (95% CI)
Cisapride without contraindicated drugs	57 650	35 690	520	1.00	1.00
Cisapride with contraindicated drugs	1873	1175	44	2.63 (1.91 to 3.63)	1.73 (1.25 to 2.40)
recent users ($0 \leq t \leq 30$ days)	8	13	10	118.94 (63.5 to 222.6)	14.08 (7.41 to 26.76)
past users ($30 < t$ days)	1865	1162	34	1.99 (1.37 to 2.87)	1.33 (0.92 to 1.93)

^aAdjusted for age, sex, respiratory system disease, digestive system disease, heart disease, and neoplasms using Cox's regression model with time-dependent covariate.

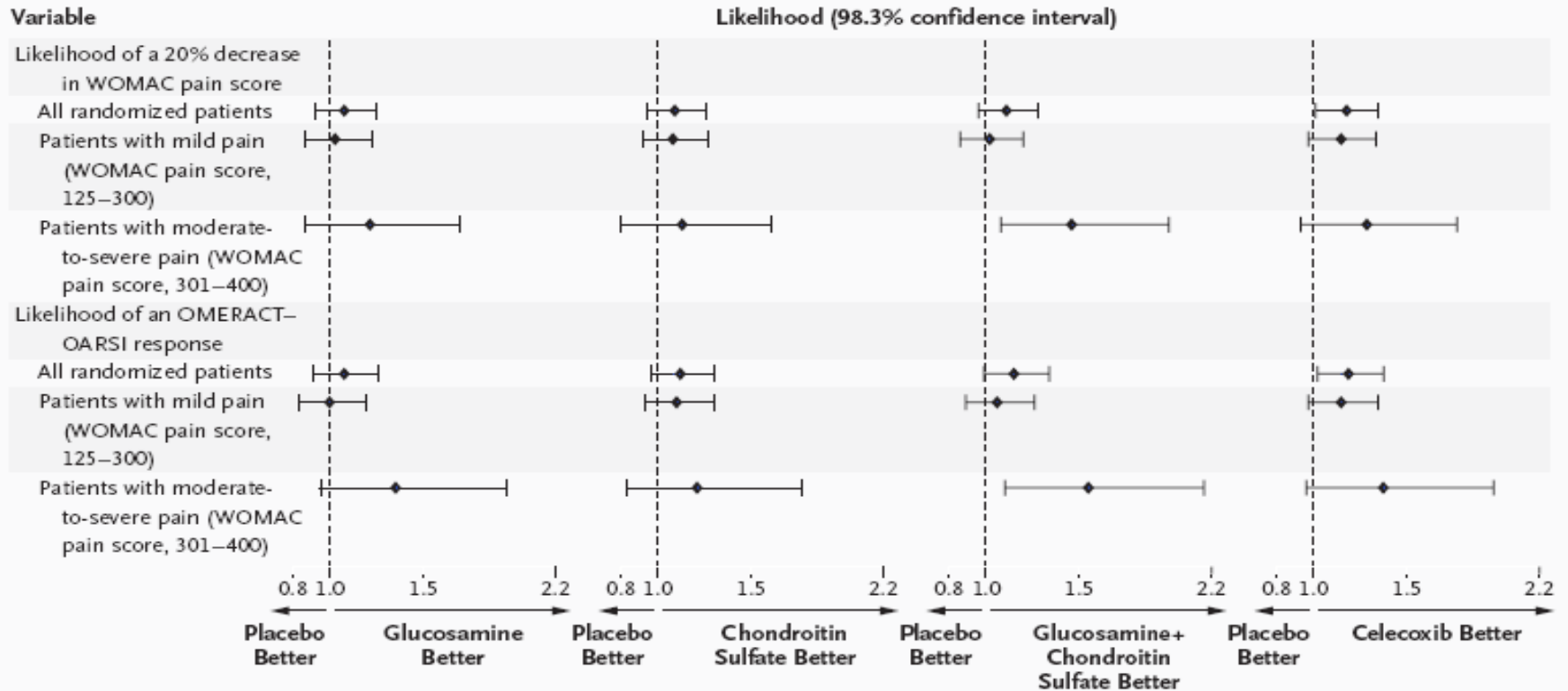


Figure 2. Pairwise Comparisons of the Overall Likelihood of a Response.

Scores for the pain subscale of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) can range from 0 to 500, with higher scores indicating more pain. A response according to the guidelines of the Outcome Measures in Rheumatology Clinical Trials and Osteoarthritis Research Society International (OMERACT-OARSI) was classified as an improvement in function or pain of at least 50 percent and a decrease of at least 20 mm on the visual-analogue scale for pain or function or the occurrence of at least two of the following: a decrease in pain of at least 20 percent and at least 10 mm on the visual-analogue scale; an improvement in function of at least 20 percent and a decrease of at least 10 mm on the visual-analogue scale; and an increase in the patient's global assessment score by at least 20 percent and at least 10 mm on the visual-analogue scale.



연구결과 해석과 결론

❖ 연구결과 해석

- 잠재적 비뚤림(bias)에 대한 충분한 고찰!

❖ 결론

- 연구자가 해당 연구를 통하여 얻은 소견과 의의
- 부족한 부분: 향후 수행이 필요한 추가적인 연구방향 제안

A close-up photograph of a hand in a purple sleeve holding a silver pen, positioned as if about to write on a document. This image is part of a decorative header bar.

Thank you!

