Statistical review from the editor's point of view
Statistical Considerations in Research Paper Writing

Jun. 2024

Dae Ryong Kang, Ph.D.

Department of Precision Medicine & Biostatistics

Yonsei University, Wonju College of Medicine

***Blaise Pascal* (1623-1662) :**

$$P(A) = 1 - P(\bar{A})$$

***Thomas Bayes* (1702-1761) :**

$$P(B_k / A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(A/B_k) \cdot P(B_k)}{\sum_{i=1}^{m} P(A/B_k) \cdot P(B_k)}$$

***Francis Galton* (1822-1911) :**

Galton's **"law of universal regression"**
who used it to characterize a tendency towards
mediocrity observed in the offspring of parent seeds.

***Erasmus Darwin* (1731-1802)**

***Charles Darwin* (1809-1882)**

***Florence Nightingale* (1820-1910)**

***Gregor Mendel* (1822-1884)**

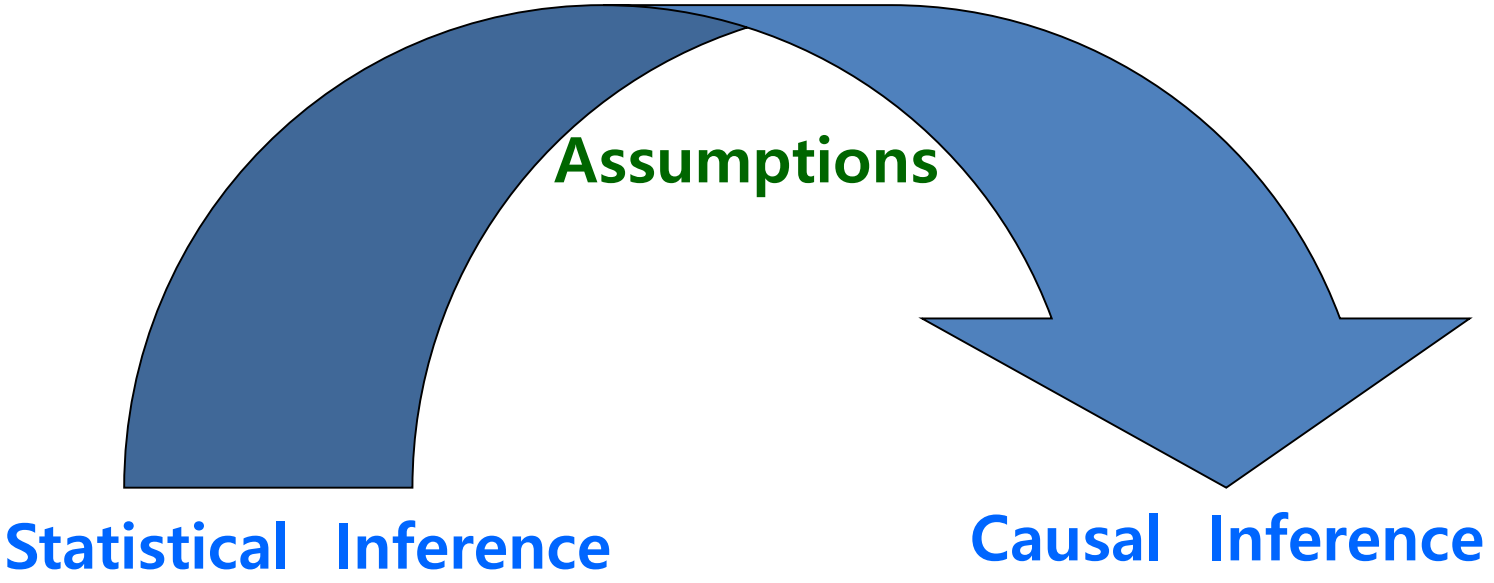
***Karl Pearson* (1857-1936) : Correlation Analysis**

***Ronald Aylmer Fisher* (1890-1962) : Design of Experiment, ANOVA, Exact test**

***Jerzy Neyman* (1894-1981) : Test of Statistical Hypothesis (H_0 vs. H_1)**

Start of *"Modern Statistics"*

What do you want to know?



What is Statistics ? It is *"learning from data"* .

Data knowledge **consists of the following three elements.**

1. **Creating valid and reliable data**
2. **Ability to analyze data**
3. **Strategic use of analysis results**

What makes it difficult for Medical Research?

① the research target is 'human'
- ethical problems
- limit of study design
- problems caused from the limit of study design

② distortion of research results occurs when we have no enough time
- need for comparing several analytic results
- lack of reflections in the discussion part

③ data noise
data incomplete
- outliers
- missing value
- there is no data without 'noise'

④ when we use inappropriate statistical methods in data analysis

Study design Study design Study design Study design **Data collection** **Data analysis** Interpretation & Discussion

EXCEL MSACCESS sas IBM SPSS R POWERPNT Hwp

1. Descriptive Statistics :

⬇ we work on **'data cleaning'** while calculating **DS** of each observed variable
e.g., n, missing value, checking outlier, category regrouping, …

2. Statistical Testing (検定) :

under significant level $\alpha = 5\%$

⬇ Statistical inference
(1:1, 1:k, sub-group)

H_0 H_A

3. Statistical Interpretation :

decision making with 'p-value'

⬇ (highly) significant, limit significant, borderline significant, not significant
(difference, association, correlation, influence with adjustment, …)

4. Interpretation through 'Medicine' or 'Public Health'

Categories of statistical procedures used to assess the statistical content in the articles

Statistical Contents	Recommended Statistical Analysis Method
Case reports, Clinical studies, Analysis of treatment result, etc.	No statistical method or Descriptive study
Evaluating the performance of the model, Setting the cut-off value (or reference value)	Sensitivity, Specificity, AUC ROC curve
Comparison of means between two paired groups	**Paired t-test, Wilcoxon signed rank test ***
Comparison of means between two independent groups	**t-test, Wilcoxon rank sum test *, Mann-Whitney U test ***
Comparison of means in three or more independent groups (or comparison between groups)	**ANOVA (with multiple comparison), Kruskal-Wallis test ***
Comparison of means measured more than three times in the same person	**Repeated measures of ANOVA, Friedman test ***
Comparison of frequency in two or more groups	Chi-squared test *, Fisher's exact test *
Comparison of frequency measured repeatedly for the same person	McNemar's test *
Correlation analysis between two continuous variables	**Pearson's correlation, Spearman's rho ***
Analysis of the relationship between dependent and independent variables	**Simple linear regression, Multiple (logistic) regression**
Esitmating survival rate, Comparing survival rate Regression of Survival data	Life table, Kaplan-Meier method Log-rank test, Cox's proportional hazard model (**HR**)
Analysis of Epidemiological statistics	Incidence, Prevalence, Risk ratio (**RR**), Odds ratio (**OR**)

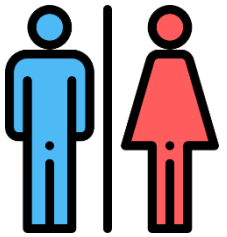
* **Non-parametric method**

Source : Emerson JD, Colditz GA. Use of Statistical Analysis in The New England Journal of Medicine. ***N Engl J Med*** 1983; 309: 709-713.

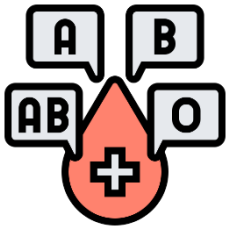
Classification of variables

❖ Categorical Variables

① Nominal scale



Gender

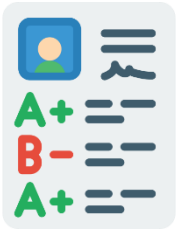


Blood type

Nationality

② Ordinal scale

Risk level



Grade

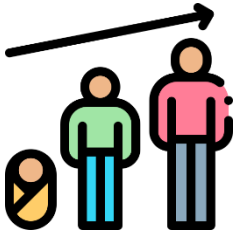
❖ Continuous Variables

① Interval scale

Temperature

IQ/EQ

② Ratio scale



Age

Height

Weight

Tips for data entry

❖ Categorical Variables

- Give each category a numeric code, and define what each number means.

Code	Blood type
1	A
2	B
3	AB
4	O

- For binary data, use 0 and 1.

Code	Event	Sex
0	No	Female
1	Yes	Male

❖ Continuous Variables

- Enter the data the **same** as measured.
- **Units** of measurement should be consistent.

❖ Other considerations

- Assign identification number(**ID**) for additional data combination.

ID	Date	Sex	Blood type	Nationality
1	20221208	.	2	1
2	20221209	0	4	3

- The date and time should be written in a **unified format (yyyymmdd)**.

ID	Date	Sex	Blood type	Nationality
1	**20221208**	.	2	1
2	**20221209**	0	4	3

- Missing values should be entered as default values (**.** or **blank**).

ID	Date	Sex	Blood type	Nationality
1	20221208	.	2	1
2	20221209	**0**	4	

Data errors checking

❖ Categorical Variables

- Frequency tables can be used to identify errors.

Values		Frequency	%
0	Female	25	25.0
1	Male	74	74.0
11		**1**	**1.0**
Total		100	100.0

❖ Continuous Variables

- Hard to find typos because of the problems of decimal points, etc.
  ➔ checking range(Min ~ Max) is effective

Descriptive statistics – variable : height of female	
Mean	171.9
Min	154
Max	**289**

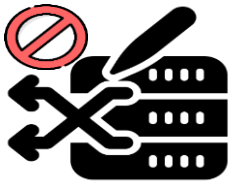
❖ Data correction

✓ Double entry of data is recommended

✓ Logical check of date

✓ Only when there is clear evidence that the data was entered incorrectly.

✓ Even if it is an incorrectly entered value, it is dangerous for researchers to arbitrarily change it to a specific value (➔ data manipulation).



Outliers, Extreme Outliers

❖ Why should we deal with outliers?

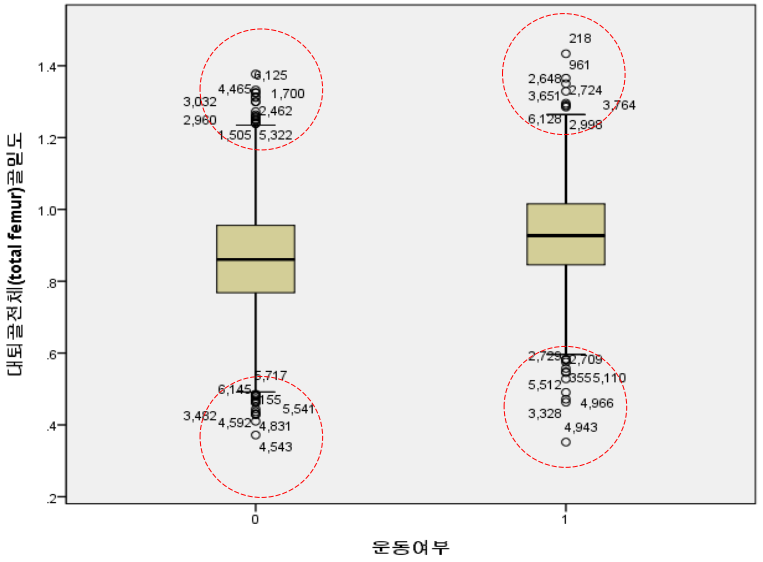
✓ They may be actual values or may be typos.

✓ Outliers must be reviewed as essential since they can seriously affect the results.

ID	Sex	Height1	Height2
1	0	154	154
2	0	162	162
3	0	171	171
4	0	154	154
5	0	159	159
6	0	163	163
7	0	164	164
8	0	168	168
9	0	166	166
10	**0**	**189**	**289**
		↓	↓
Mean		165.00	**175.00**
S.D.		10.08	**40.43**

❖ How can we identify/handle outliers?

✓ Outliers can be checked through range check (min/max), graphical methods (box plot).



✓ It should be avoided to delete outliers unconditionally.

✓ It is desirable to perform "**with & without**" analyses to confirm the similarity of the results.

Normality test

❖ Why should we do the normality test?

It checks whether a given data is randomly drawn from a regular population.

Yes → **Parametric test**
e.g. Independent t-test

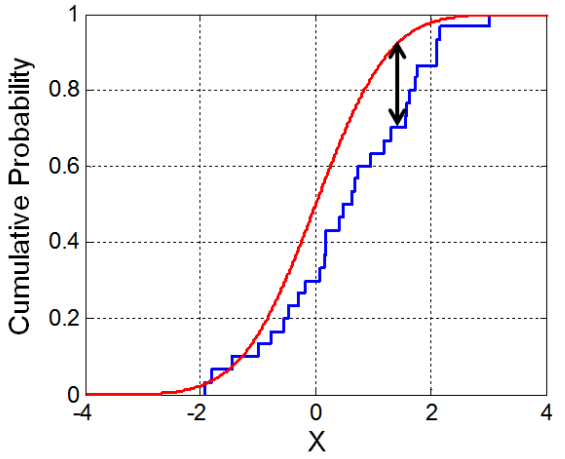
No → **Non-parametric test**
e.g. Mann-Whitney U-test

Normality test can be done by …

- ✓ Checking histogram of the data and compare it to the **normal distribution**.
- ✓ Using statistical methods.
  e.g. Kolmogorov-Smirnov test, Shapiro-Wilk test

❖ Normality test methods

① Kolmogorov-Smirnov test ← **N= 30 or 50** ↑



Red line: model CDF
Blue line: empirical CDF
Black arrow: KS statistic.

② Shapiro-Wilk test ← **N= 30 or 50** ↓

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$(a_1, \ldots, a_n) = \frac{m^{\mathsf{T}} V^{-1}}{C}$$

V is the covariance matrix of those normal order statistics

$$C = \|V^{-1}m\| = (m^{\mathsf{T}} V^{-1} V^{-1} m)^{1/2}$$

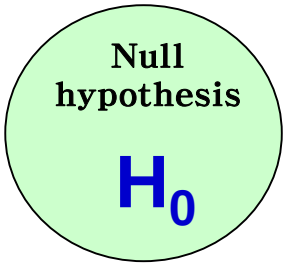
$$m = (m_1, \ldots, m_n)^{\mathsf{T}}$$

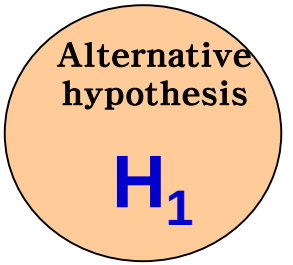
Non-parametric Statistical Analysis

- ✓ sample size is not large and the distribution of the population does not follow a normal distribution.
- ✓ based on their relative 'rank' or 'order' rather than actual values.
- ✓ It compares the median(Q2), not the mean.
- ✓ degree of variance is expressed as a 'range' or 'IQR' rather than a standard deviation(SD).

Parametric method	Non-parametric method
Paired t-test	Wilcoxon signed rank test
Independent two-sample t-test	Wilcoxon rank sum test Mann-Whitney U test
One-way ANOVA Two-way ANOVA	Kruskal-Wallis test Friedman test
Pearson's correlation	Spearman's rank correlation Kendall's tau

Errors in Hypothesis Testing





H_0 : No difference between effects of two drugs

H_1 : Not H_0

H_0: μ=120 mmHg vs. H_1: $\mu \neq$120 mmHg

Decision \ True	H_0 is True	H_0 is False
Fail to Reject H_0	☺ Ex. = Control	☹ β Type II error Ex. = Control
Reject H_0	☹ ☹ α Type I error Ex. ≠ Control	☺ ☺ Ex. ≠ Control "Power (1-β)"

Type I error = P(positive | H_0 true) = "False Positive"

Type II error = P(negative | H_0 false) = "False Negative"

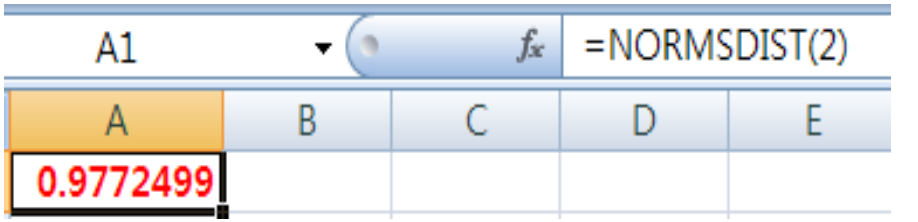
p-value vs. α-level

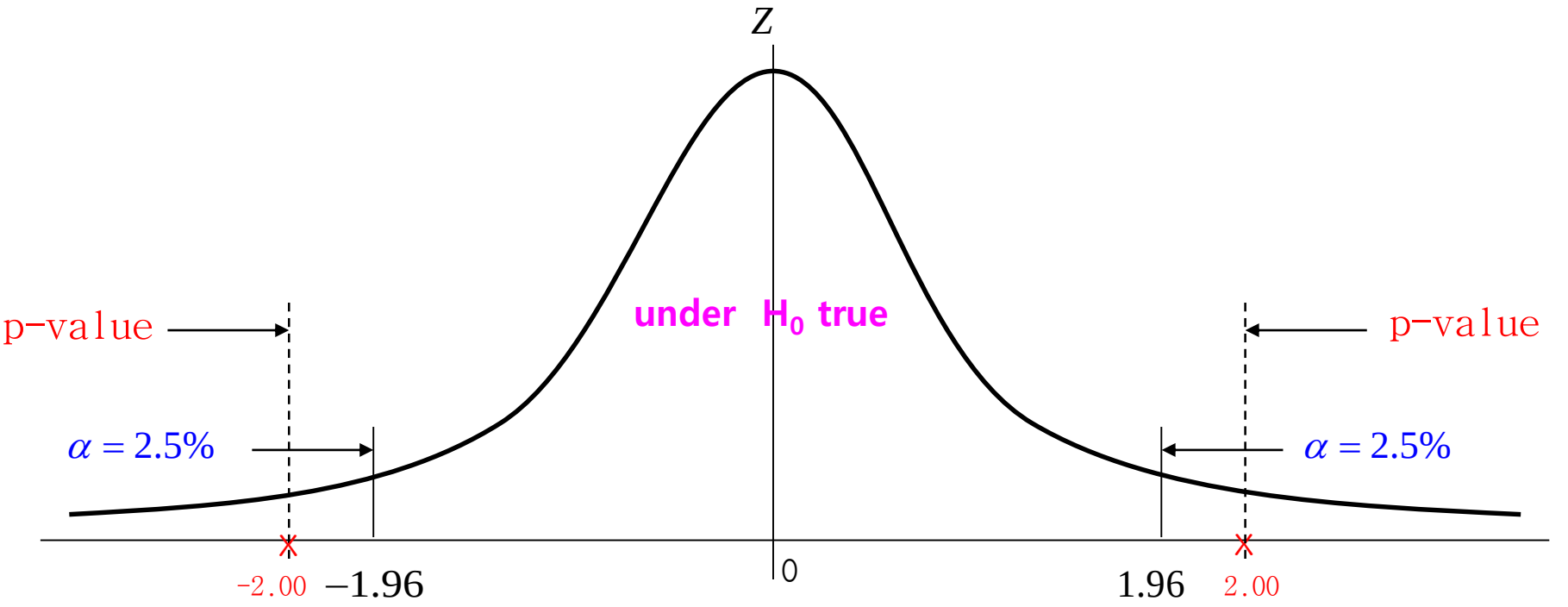
- conclusion based on a **statistical testing** is typically reported in conjunction with a *p-value*

- **p-value :** actual probability of obtaining the particular sample outcome from a population for which **H_0 is true**
- **α-level :** the risk of incurring a **type I error** that the investigator is willing to tolerate ***(significance level)***

" Rejcet H_0 and concluded that H_1 is true (accept) If p-value ≤ α=0.05 "

p-value = (1 - 0.9772499) * 2 = 0.0455





p-value ?

the probability of obtaining a result at least as extreme as the one that was actually observed, given that the H_0 is true

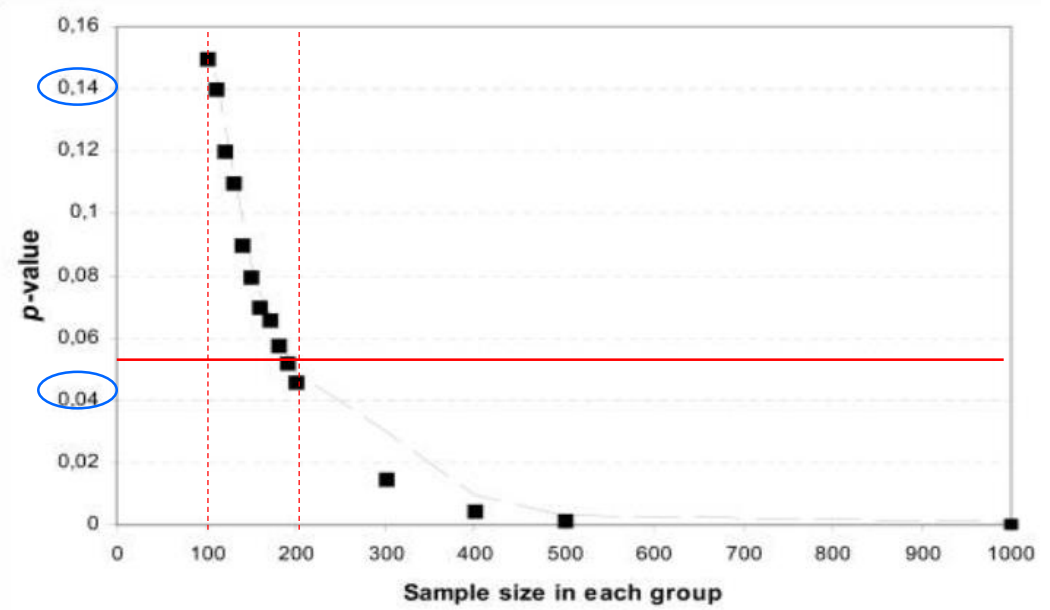
p-value in Biomedical Research

- ***Is the p-value of 0.05 an absolute criterion for determining 'significance' ?***

- ***Statistical significance interpretation based on p-value has limitation.***
  ➔ p-value is an important criterion in decision, but has some limitations.

- ***The p-value doesn't explain the degree of importance of the observed effect.***
  ➔ p-value is small does not necessarily mean that the association is strong.

- ***The p-value is closely related to the sample size.***
  ➔ Interpreting the statistical results to rely only on p-value would be difficult to prove the improvement of clinical usefulness because the target sample size is not achieved, especially when dealing with "*rare diseases*".



Source: DB Panagiotakos,
The Open Cardiovascular Medicine Journal,
2008, 2, 97-99.

Table 1. Key Questions to Ask When the Primary Outcome Is Positive.

Does a P value of <0.05 provide strong enough evidence?
What is the magnitude of the treatment benefit?
Is the primary outcome clinically important (and internally consistent)?
Are secondary outcomes supportive?
Are the principal findings consistent across important subgroups?
Is the trial large enough to be convincing?
Was the trial stopped early?
Do concerns about safety counterbalance positive efficacy?
Is the efficacy–safety balance patient-specific?
Are there flaws in trial design and conduct?
Do the findings apply to my patients?

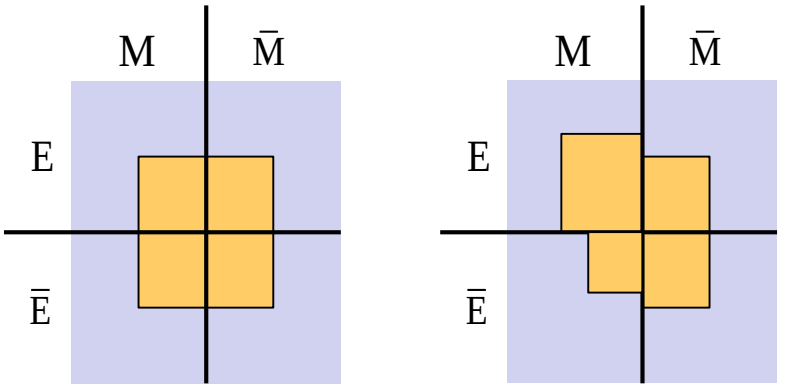
← *Stuart et al., N Engl J Med 2016;375:971-9.*

Table 1. Questions to Ask When the Primary Outcome Fails.

Is there some indication of potential benefit?
Was the trial underpowered?
Was the primary outcome appropriate (or accurately defined)?
Was the population appropriate?
Was the treatment regimen appropriate?
Were there deficiencies in trial conduct?
Is a claim of noninferiority of value?
Do subgroup findings elicit positive signals?
Do secondary outcomes reveal positive findings?
Can alternative analyses help?
Does more positive external evidence exist?
Is there a strong biologic rationale that favors the treatment?

Stuart et al., N Engl J Med 2016;375:861-70. →

at the research design stage bias



① Selection bias

- ✓ sampling frame bias : admission rate bias (*Berksonian* bias)
- ✓ non random sampling bias : detection bias
- ✓ non-converge bias : loss to follow-up bias, withdrawal bias

② Non comparability bias

- ✓ lead time bias, length bias, historical control bias

③ Sample size bias

in the data collection process information bias

① Instrument bias

② Data source bias

③ Observer bias

- ✓ **diagnostic suspicion bias**
- ✓ **exposure suspicion bias**
- ✓ **therapeutic bias** (➔ Blinding)

④ Subject bias

- ✓ **proxy respondent bias**
- ✓ **recall bias**
- ✓ **attention bias** ("Hawthorne effect")

in the process of analysis & interpretation of results bias

① **Confounding bias**

② **Analysis strategy bias**

: **missing data handling, outlier handling, unit of analysis**

③ **Post-hoc analysis bias** (← data dredging bias)

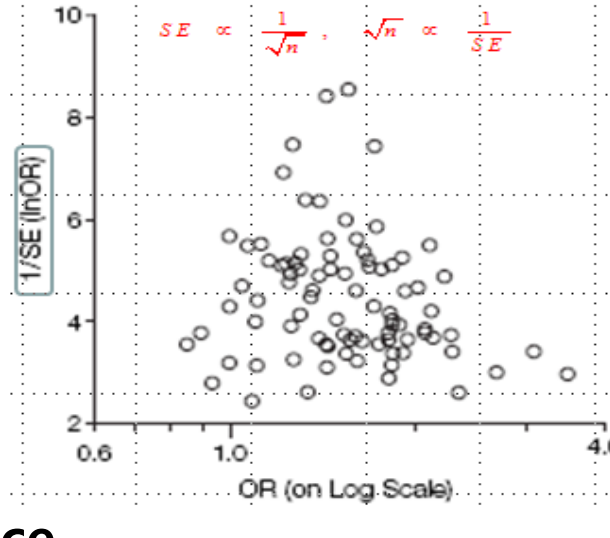
④ **Assumption bias**

⑤ **Generalization bias** (← lack of external validity)

⑥ **Significance bias**

: **statistical significance vs. biological significance**

⑦ **Publication bias** (by Funnel plot, Egger's regression asymmetry test)



- ✓ Association (연관성 聯關性)
  - Homogeneity (동질성 同質性)
  - Independency (독립성 獨立性)
- ✓ Correlation (상관성 相關性)
- ✓ Probability (개연성 蓋然性)
- ✓ Causality (인과성 因果性)

4 Major Bigdata in Bio-Healthcare

"Big Data (5V1C)"
Volume + Variety + Velocity +
Value + Veracity + Complexity
Data Visualization

토탈오믹스 Total-omics

- 01 / 전장유전체 (Whole Genome)
- 02 / 엑솜 (exome) 시퀀싱
- 03 / 타깃 시퀀싱
- 04 / 전사체 (Transcriptome)
- 05 / 후성유전체 (Epigenome))
- 06 / 단백질체 (Proteomics)
- 07 / 대사체 (metabolomics)
- 08 / 미생물 정보 (microbiome)

GENOME

Multi-OMICS
KoGES

병원/개인 진료정보
EMR/EHR/PHR

h·well 국민건강보험 National Health Insurance Service
건강보험심사평가원
질병관리본부
기상청
식품의약품안전처

Lifelog Data : 일상생활에서 기록 및 저장되는 모든 정보를 의미
Wearable technology
Mobile devices
PGHD(Patient-Generated Health Data)

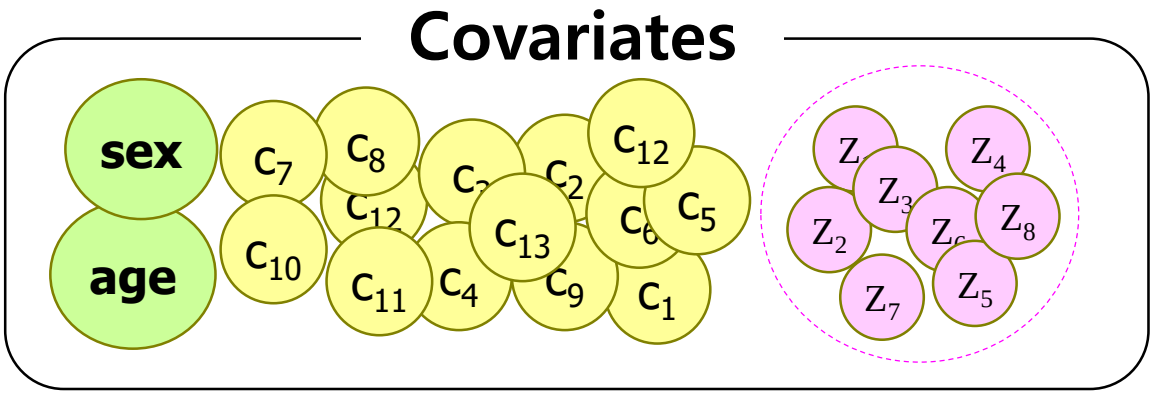
Data

Information ➡ Knowledge ➡ Theory & Expertise

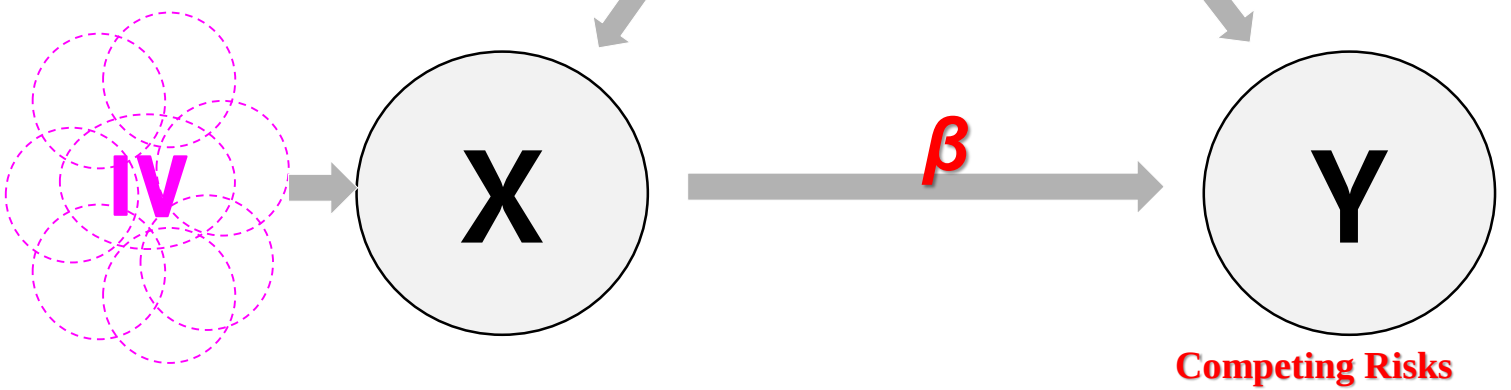
"구슬이 서말이라도 꿰어야 보배다." 그러나 애초에 그 구슬이 참 보배여야 한다.

Statistical Methods for Causal Inference

- confounders variables
- unmeasured/unknown confounders
- stratification variables
- intermediate variables
- effect modifier / interaction effect



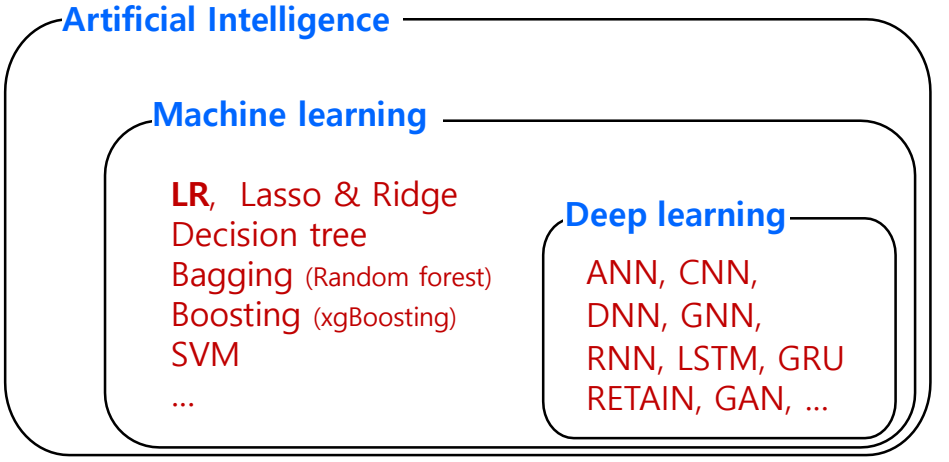
+ 'Time' correction?

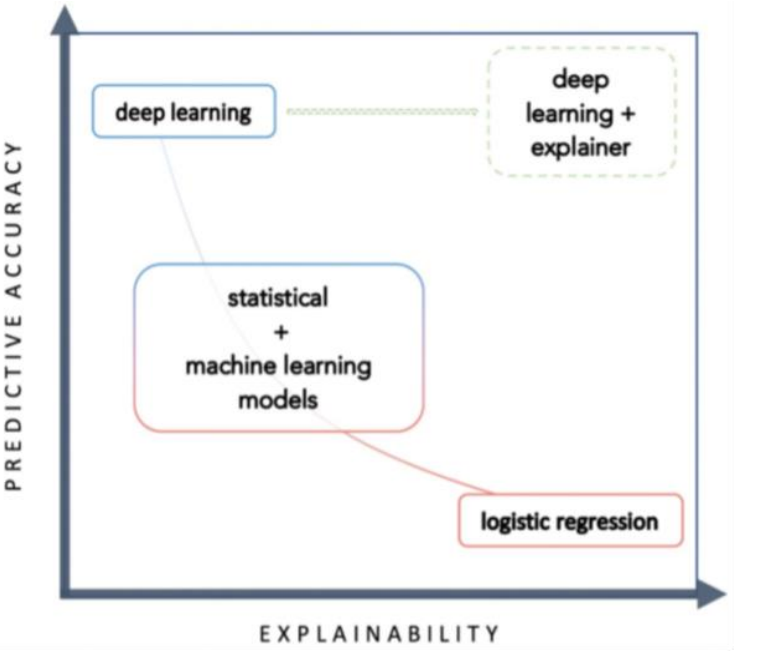


- Multiple Regression Analysis
- Logistic Regression Analysis
- Poisson Regression Analysis
- Cox's PHM
- Linear Mixed Model (**LMM**)
- Generalized Estimating Equation (**GEE**)

LR - ML - DL

- There is no bright line between machine learning models and traditional statistical models
- **Deep learning** is well suited to learn from the complex and heterogeneous kinds of data that are generated from modern clinical care, such as medical notes entered by physician, medical images, continuous monitoring data from sensors, and genomic data to help make medically relevant predictions.





Python3.8 / BioPython / Anaconda3 / Jupyter

➔ feature 'importance' ranking

FDA 21st Century Cures ACT 만장일치로 하원 통과 (2016.12. 13.)

- FDA는 RWE를 활용한 허가심사체계의 기반을 마련하기 위해, RWD를 RWE로 변환하는 작업에 참여하고 있음.
- 「FDA 21st Century Cures ACT」 법안 도입으로
  - FDA는 승인된 의약품의 적응증 추가나 시판 후 연구에 활용될 RWE의 잠재적인 사용을 평가할 프로그램을 만들도록 요구
  - FDA는 2018년 말까지 이 프로그램을 위한 기틀을 마련해야함.
  - FDA는 2021년 말까지 가이드라인 초안 발행
  - RWE 허가심사 반영 핵심 두 가지
    1) 새로운 적응증의 추가 /신약허가 대조군 활용
    2) 시판 후 안전성 요건의 강화

The US FDA에서 Advancing RWE Program을 발표함 (2022.10.22.)

9. Attributes of study design: objectives; design architecture (e.g., randomized trial with pragmatic elements, externally controlled trial, observational cohort study) with a schematic representation; eligibility criteria; covariates of interest; primary and key secondary endpoints; treatment of interest, comparator, and concomitant therapies

10. Potential data sources: category (e.g., EHR, medical claims, registries, and/or other) and description; data reliability and relevance; validation, timing, and completeness of key data elements; linkage to other data sources; additional data collection.

11. Anticipated analysis plan: approximate sample size; analytic plan for primary and key secondary endpoints; approach to confounding factors; definition of follow-up period; handling of intercurrent events, missing or misclassified data, and multiplicity.

12. Miscellaneous considerations: pre-specification of study design and conduct; availability and FDA access to patient-level data; approach to human subject protection.

실사용데이터

Real World Data

Real-World Data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.

- Electronic Health Records (EHRs)
- Medical claims and billing data
- Data from product and disease registries
- patient-generated data, including from in-home-use settings
- Data gathered from other sources that can inform on health status

실사용근거

Real World Evidence

Real-World Evidence (RWE) is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.

- Not limited to Randomized trials (e.g.,large simple trials, pragmatic clinical trials)
- Externally controlled trials
- Observational studies (prospective or retrospective)

'외부대조군' 이란? 동일한 무작위배정임상시험(RCT)에 참여하지 않은 환자들로 구성된 대조군을 의미하며, 전자의료기록 및 과거 임상시험 데이터 등을 기반으로 구축된다.

RCT + RWD/RWE = 글로벌 임상연구 강국

❖ 실용적 RCT의 최대 장점은 '대규모' 환자를 대상으로 '단기간' 내에 연구를 완료할 수 있다는 것이다. 게다가 연구에 큰 비용이 들지 않는다. (TASTE 연구, N Engl J Med 2013;369:1587-1597).

❖ 실용적 RCT는 다양하고, 대표성을 띠고, 이질적인 환경과 인구집단으로 부터 환자를 등록하고, 새로운 전략을 현재 수용된 표준과 비교하고, 추적관찰 의료결과 자료수집에 초점을 맞춤. (JAMA 2018;320(2):137-138.)

➔ 우리나라는 전국민 단일건강보험 체계로, NHIS, HIRA에 전국민들의 의료데이터가 모이기 때문 실용적 RCT가 우리나라에 **최적화된 연구디자인**이라고 평가한다.

➔ 우리나라에서는 실용적 RCT를 하기 위한 걸음마조차 떼지 못한 실정이다. 가장 큰 걸림돌은 '**연구참여 동의서**' 취득과 '**IRB 심의**' 통과이다. 아직 실용적 RCT의 필요성을 인식하고 있는 연구자 수가 적어 연구가 활성화되지 않아, IRB 심의에서 **실용적 RCT에 대한 인식이 크지 않다.**

스마트 임상시험 신기술개발 연구사업

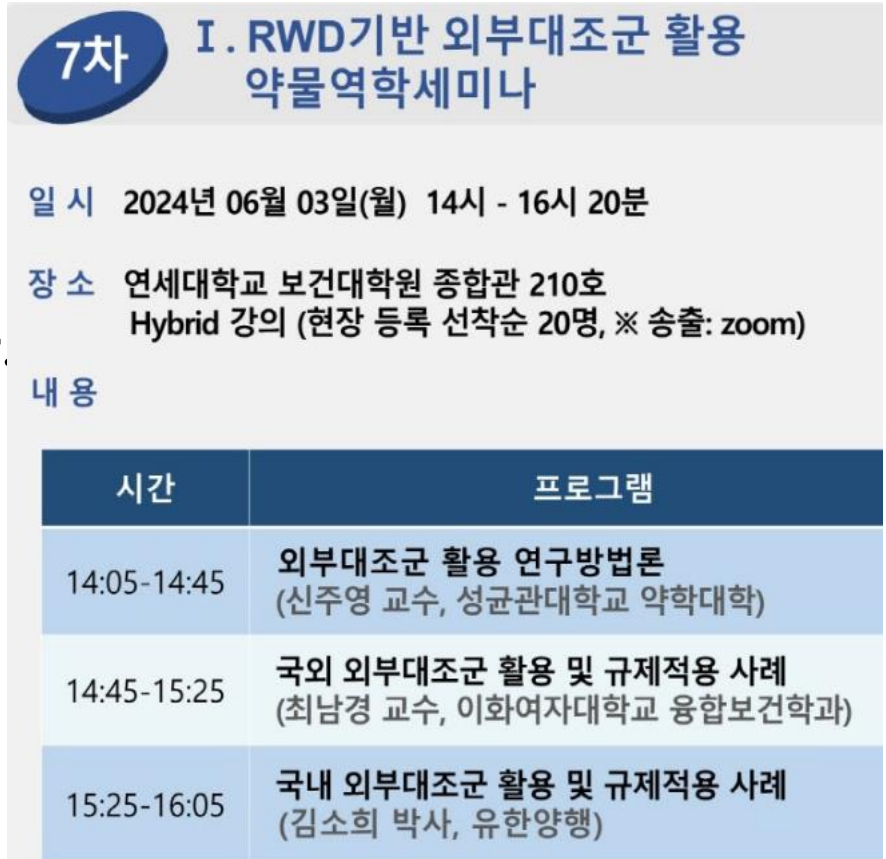
카카오헬스케어 + C&R리서치 + 경희의료원 (2023.7)

➔ '외부대조군' 데이터 기반 글로벌 임상시험 사업화

'외부대조군' 이란?

동일한 무작위배정임상시험(RCT)에 참여하지 않은 환자들로 구성된 대조군을 의미하며, 전자의료기록 및 과거 임상시험 데이터 등을 기반으로 구축된다.

➔ 특히 '외부대조군'은 희귀질환 및 난치질환을 대상으로 하는 무작위배정 임상시험 RCT의 비윤리성 및 대상자 모집의 어려움을 극복하고 치료제 개발 지연문제를 해결하고자 활용되고 있다.



7차 Ⅰ. RWD기반 외부대조군 활용 약물역학세미나

일 시 2024년 06월 03일(월) 14시 - 16시 20분

장 소 연세대학교 보건대학원 종합관 210호
Hybrid 강의 (현장 등록 선착순 20명, ※ 송출: zoom)

내 용

시간	프로그램
14:05-14:45	외부대조군 활용 연구방법론 (신주영 교수, 성균관대학교 약학대학)
14:45-15:25	국외 외부대조군 활용 및 규제적용 사례 (최남경 교수, 이화여자대학교 융합보건학과)
15:25-16:05	국내 외부대조군 활용 및 규제적용 사례 (김소희 박사, 유한양행)

Thank you for listening.

Q & A

iNHCBCR 국민건강빅데이터임상연구소
National Health BigData Clinical Research Institute