

# Biostatistics of RCT and Observational Studies

안형진 교수  
고려대학교 의과대학  
의학통계학교실

# 서론

- 통계학의 중요한 역할
  - 궁극적으로 연구가 비교할만하고(comparable) 일반화할 만한(generalizable)한지 증명
  - 연구계획 및 설계가 매우 중요
- 연구는 인과관계 추론에서 발생할 수 있는 여러 가지 편향(bias)을 최소화할 수 있도록 설계해야 함.
  - 전향적 무작위 중재연구 (best)
  - 비무작위 연구 (bias 통제 필요)
  - 관찰연구 (전향적 코호트연구, 사례-대조 연구, bias 통제 필요, 일반화가 좋을 수 있음)
- 잠재적 교란변수(potential confounders)를 미리 수집

# 서론

- 편향 (Bias)
  - 연구결과와 실제현상 사이에 구조적인 차이 (systematic difference)가 존재하면 편향(bias)이 발생되었다고 한다
    - 조사자 편향 (observer bias) 및 평가편향 (assessment bias) – 임상시험(눈가림)
    - 교란편향 (confounding bias) – 관찰연구 및 비 무작위배정
    - 선택편향 (selection bias) – 관찰연구
    - 할당편향 (allocation bias) – 임상시험(무작위배정)
    - 정보편향 (information bias) - 측정오차
    - 출판편향 (publication bias) – Systematic Review
    - 기억편향 (recall bias) – 사례-대조 연구
    - 건강인 진입효과 (healthy entrant effect) – 코호트 연구

# 서론

- 연구설계의 중요성
  - 잘못된 연구설계 → 부적절한 결론도출
  - 잘못된 연구설계 → 어떤 통계적 방법도 설계상의 문제를 해결할 수는 없다.
  - “... a poor design cannot be salvaged by good statistics.”
- Good Work → Team Work
  - Get biostatisticians on board in the beginning of the study.

# 연구설계의 종류

## ■ 실험연구 (Experimental, Interventional Study)

- 외부요인(extraneous factor)들을 통제하면서 연구에 관련된 요인을 조작
- Randomized Controlled Trials (RCT)은 실험연구의 한 예 (다른 위험요인들의 직접적 통제는 불가능).
- 일반적으로 연구자가 결과에 영향을 미칠지도 모르는 다른 요인들을 통제할 수 있으므로 연구가설을 증명 (또는 인과관계의 추론)할 수 있는 가장 확실한 연구설계라 할 수 있다..
- 윤리문제 등 여러 가지 제약으로 특히 사람을 대상으로 하는 연구에서는 불가능한 경우가 많다.

# 연구설계의 종류

- RCT의 장점
  - 엄격하고 근거중심
  - 객관적이고 독립적
  - 할당편향(교란편향)의 최소화
  - 알려지거나 알려지지 않은 기저변수의 보정
  - 인과관계를 조사할 가능성이 높음
- RCT의 단점
  - 대조군의 선택이 어려움
  - 비용과 시간
  - 윤리적인 문제
  - 일반화의 문제
  - 결과변수의 제한

# 연구설계의 종류

- 관찰연구 (Observational Study)
  - 실험에 미치는 요인을 통제할 수 없으며 단지 관찰만 가능한 연구
  - 예: Cohort Study, Case-Control Study, Cross-Sectional Study (표본조사)
  - 역학연구(epidemiological study): 일반 인구집단을 대상으로 관심요인과 질병과의 관계를 평가하는 연구이며 관찰연구에 속한다.

# 연구설계의 종류

- 관찰연구의 장점
  - 유연성
  - 일반적으로 윤리적
  - 상대적으로 큰 표본수
  - 일반화의 가능성
- 관찰연구의 단점
  - 대조군 선택의 어려움
  - 선택편향 (교란편향)의 발생
  - 인과관계를 밝히기 어려움
  - 군간 표본수의 불균형

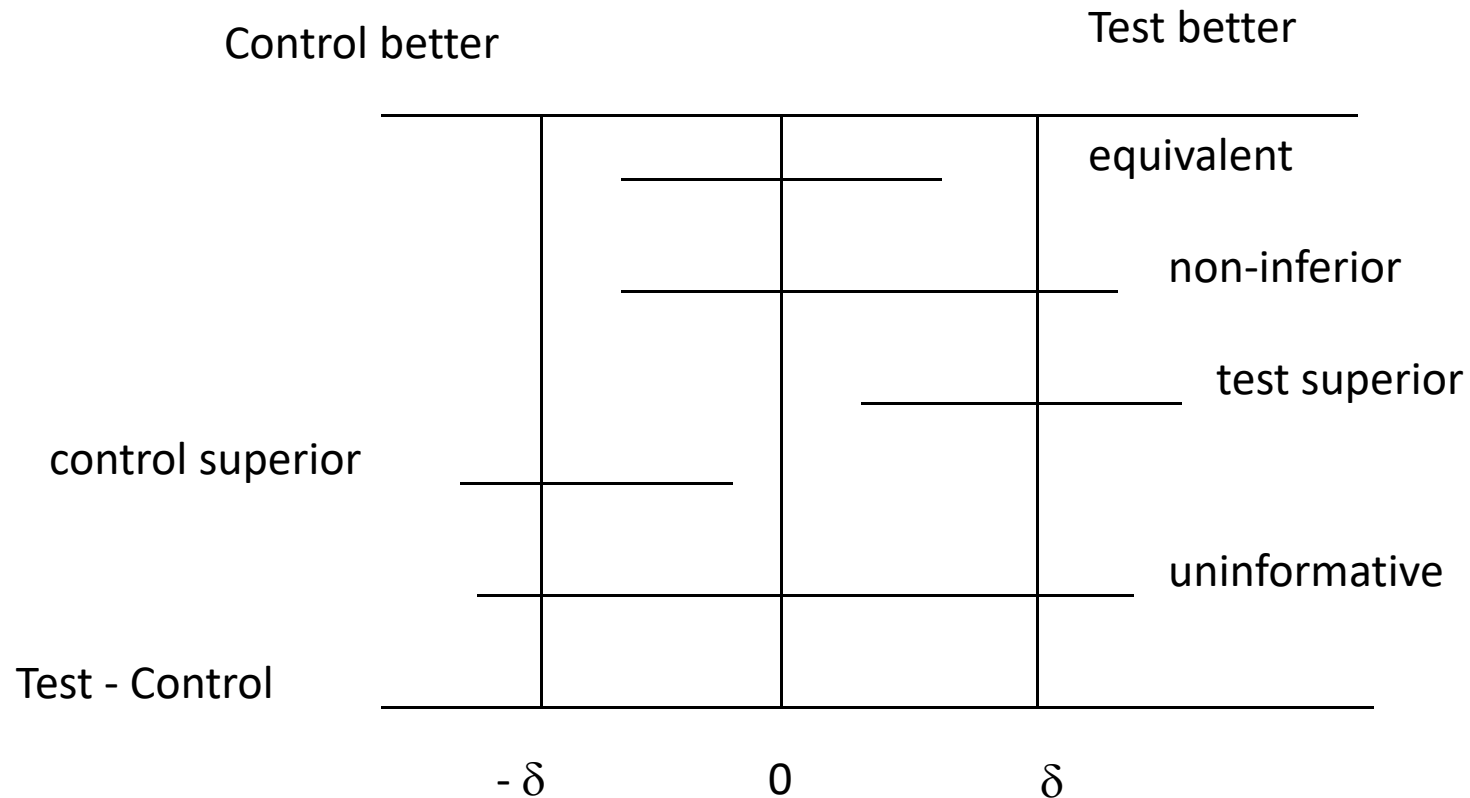


# 비교검정의 종류 (RCT)

- 우월성 검정 (Superiority Test)
  - 연구목표가 대조군에 비하여 처리군의 효과가 더욱 뛰어남을 보이려는 경우에 실시하는 검정
  - 일반적으로 시험용 의료기기와 Sham 의료기기의 비교를 하는 경우에 적용한다.
  - 전통적인 통계적 검정방법
- 비열등성 검정 (Noninferiority Test)
  - 연구목표가 처리군의 효과가 대조군보다 열등하지 않음을 보이려는 경우에 실시하는 검정
  - 윤리적인 문제 (즉, 효과가 있는 기기가 이미 존재함에도 불구하고 Sham을 사용하는 것은 윤리적이지 못함) 또는 generic 기기의 효과를 입증하려고 하는 경우에 비열등성 검정을 실시함.

# 비교검정의 종류 (RCT)

- 이러한 검정, 특히 비열등성 검정은 신뢰구간을 이용하여 의사결정을 함.



# 비열등성 검정

- 미리 정해진 임상적으로 무시할 수 있다고 판단되는 정도의 차이를  $\delta$  (비열등성 한계치: non-inferior margin)라고 할 때, 비열등성은 활성대조군과 처리군의 효과 차이에 대한 신뢰구간이  $\pm\delta$ 의 범위 안에 포함되어 있는 지에 대한 평가를 통하여 검정을 실시
- 비열등성은 신뢰구간의 단측을 확인하여 평가함.
- 단측을 사용하므로 5% 대신 2.5%를 유의수준으로 사용함, 즉, 97.5% 단측 신뢰구간을 이용함.

# 비열등성 검정

- 비열등성 검정을 수행하기 위해서는 임상적으로 무시할 만한 차이, 즉 한계점에 대한 결정이 연구계획 단계에서 정해야 함.
- 한계점은 통계적인 측면뿐만 아니라 임상적인 측면에서 동등성 및 비열등성 검정의 성패를 결정짓는 중요한 요인이 됨.
- 한계점을 정하는 일반적인 규칙은 정립된 것이 없지만 “sham군과 활성대조군의 효과 차이의 절반보다 작아야 한다”는 일반적인 규칙이 존재함
- 하지만 sham의 효과가 작은 유효성 변수를 고려할 경우 치료효과와 sham효과 간의 차이가 크게 벌어져 있기 때문에 이 원칙을 그대로 적용하기에는 무리가 있음.

# 비열등성 검정

- 활성 대조군의 선정에 있어서도 주의를 기울여야 함.
- 현재 임상에서 널리 사용되고 있고 표준처치라고 고려되는 의료기기 또는 진단방법을 기준으로 삼아야 함.
- 즉, 이전 연구가 잘 정립되어 있어 선행 우월성 검정의 자료들이 충분한 것을 대조군으로 설정하여야 함.

# 분석군

- 무작위 배정(**Randomization**)은 처치군간 비교성을 보장한다.
- 무작위 배정과 시험종료 사이에 시험 대상자가 예기치 않은 이유로 무작위 배정된 처치를 따르지 않거나 다른 처치로 바꾸거나 연구에서 중도 탈락(drop-out)하는 경우가 발생한다.
- 이런 시험 대상자를 제외하고 분석하는 경우 **bias**가 발생할 수 있다.

# 분석군

- **Possible Solutions**
  - **Intention-to-treat (ITT)**
  - **Modified intention-to-treat (mITT) or Full analysis set (FAS)**
  - **Per-protocol (PP)**

# ITT

- **Intent-to-treat** 으로도 알려져 있다.
- ITT의 원칙( **principles of ITT** ): 연구참여자를 무작위 배정된 대로 비교 (**as randomized**).
- 처치(therapy)의 완전한 잠재적 이익을 포착하지 못하므로 ITT가 최적의 방법이 아닌 것으로 생각될 수도 있으나 ITT는 몇 가지 장점들이 있다.



# ITT

- 장점

- ITT 분석은 무작위배정의 강점을 보유 (즉, *bias minimization*)
- ITT는 PP보다 실제(real-life)에서 일어날 수 있는 상황을 더 고려
- ITT는 연구의 어떤 시점에서든 모든 연구참여자의 정보를 이용 → 중간분석을 가능하게 함.
- ITT분석은 처치에 있어 실제적인 정보를 제공

# ITT

- ITT는 처치에 순응한 개체, 비순응 개체, 처치를 바꾼 개체 모두 포함
- ITT는 PP처럼 처치의 최대의 잠재적 효과를 보려는데 목적이 있지 않음.
- ITT는 PP의 결과보다 그 효과가 통계적으로 유의하지 않거나 효과의 크기가 작은 경우가 많음.
- 결측자료의 처리가 쉽지 않음.

# Practice of ITT

- **Full Analysis Set (FAS)**

- 실제에서, ITT의 원칙을 그대로 따르는 것은 어려움이 있기 때문에 regulators는 어느 정도의 타협을 허용함
- FAS는 ITT의 원칙을 최대한 유지하여야 함.

- Bias없이 개체를 제외할 수 있는 가능한 경우의 예

- 포함/제외 기준에 맞지 않는 피험자
- 시험약/대조약을 한번도 복용하지 않은 피험자
- 베이스라인 측정값 이후 측정값이 없는 피험자

# Practice of ITT

- **Full analysis set (FAS)**이라는 용어는 ITT 원칙과 ITT의 실제 적용을 구분하기 위하여 사용함.
- 하지만 많은 연구자들은 ITT의 실제 적용도 ITT라고 사용하는 경우가 흔함.
- 어떤 연구자들이나 규제기관은 FAS 대신 **modified ITT (mITT)** 라는 용어를 사용하기도 함.

# Per-protocol

- FAS 분석군과 마찬가지로 PP 분석군도 프로토콜에서 분명히 명시하여야 함.
- PP 분석은 bias된 결과를 가져올 수 있으며 처치 효과가 과대평가될 수 있음.
- 이런 이유로 우월성 시험에서는 PP군은 일반적으로 부분석 (secondary analysis)에 사용됨.

## Frequently used statistical methods for independent continuous response

Situation	Statistical Methods
One sample	<ol style="list-style-type: none"> <li>1. Parametric: One sample t-test (<b>PROC TTEST</b>)</li> <li>2. Non-parametric: Wilcoxon signed rank test (<b>PROC NPAR1WAY</b>)</li> </ol>
Independent two-sample comparison	<ol style="list-style-type: none"> <li>1. Parametric: Independent two-sample t-test (<b>PROC TTEST</b>)</li> <li>2. Non-parametric: Wilcoxon rank sum test (A.K.A. Mann-Whitney U test) (<b>PROC NPAR1WAY</b>)</li> </ol>
Three or more group Comparison	<ol style="list-style-type: none"> <li>1. Parametric: analysis of variance(ANOVA) with multiple comparison (<b>PROC ANOVA or PROC GLM</b>)</li> <li>2. Non-parametric: Kruskal-Wallis test (<b>PROC NPAR1WAY</b>)</li> </ol>
Relationship with continuous predictors	<ol style="list-style-type: none"> <li>1. One predictor: simple linear regression (<b>PROC REG or PROC GLM</b>)</li> <li>2. Multiple predictors: multiple linear regression (<b>PROC REG or PROC GLM</b>)</li> </ol>
Relationship with continuous and categorical predictors	<ol style="list-style-type: none"> <li>1. Analysis of covariance (ANCOVA: <b>PROC GLM</b>) or</li> <li>2. General linear models (GLM: <b>PROC GLM</b>)</li> </ol>

## Frequently used statistical methods for correlated continuous response

Situation	Statistical Methods
Paired data	<ol style="list-style-type: none"><li data-bbox="696 703 1626 751">1. Parametric: paired t-test (<b>PROC TTEST</b>)</li><li data-bbox="696 815 1957 927">2. Non-parametric: Wilcoxon signed rank test on differences (<b>PROC NPAR1WAY, PROC UNIVARIATE</b>)</li></ol>
Repeated measures	<ol style="list-style-type: none"><li data-bbox="696 1023 1928 1126">1. Repeated measures ANOVA (RM-ANOVA: <b>PROC GLM with REPEATED Statement</b>) or</li><li data-bbox="696 1190 1693 1238">2. Linear mixed model (LMM: <b>PROC MIXED</b>)</li></ol>

## Frequently used statistical methods for independent categorical response

Situation	Statistical Methods
Relationship with one categorical predictor	<ol style="list-style-type: none"><li>1. Chi-square test with large sample (<b>PROC FREQ</b>)</li><li>2. Fisher's exact test with small sample (<b>PROC FREQ</b>)</li><li>3. Simple linear logistic regression (<b>PROC LOGISTIC, PROC GENMOD</b>)</li></ol>
Relationship with categorical and continuous predictor	<ol style="list-style-type: none"><li>1. Multiple logistic regression for binary response (<b>PROC LOGISTIC, PROC GENMOD</b>)</li><li>2. Multinomial or ordinal logistic regression for categorical response (more than two levels) (<b>PROC LOGISTIC</b>)</li></ol>



## Frequently used statistical methods for correlated categorical response

Situation	Statistical Methods
Paired categorical data analysis	1. McNemar's test ( <b>PROC FREQ</b> )
Repeated categorical responses	1. Marginal model using generalized estimating equation (GEE: <b>PROC GENMOD</b> ) 2. Generalized linear mixed model ( <b>GLMM: PROC GLMMIX, PROC NLMIXED</b> )

# Frequently used statistical methods

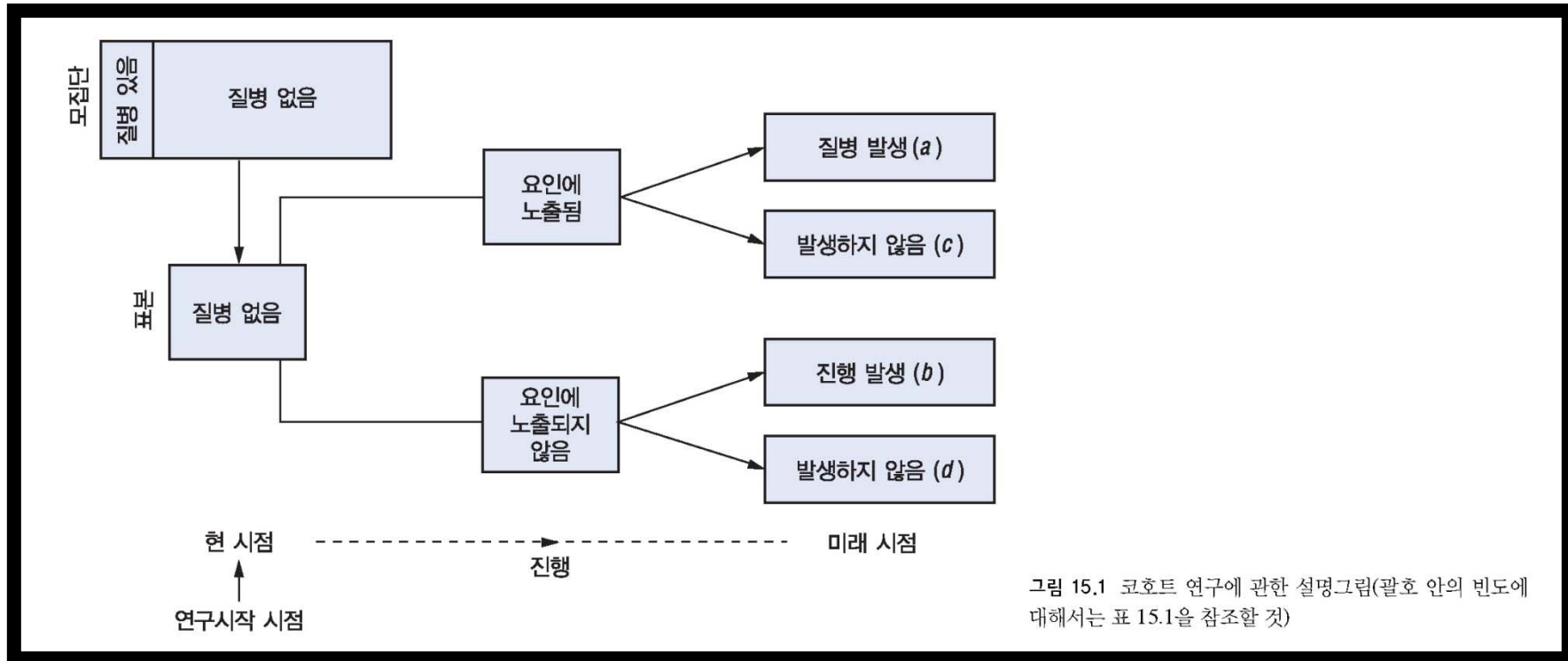
## Other methods

Situation	Statistical Methods
Correlation between two continuous variables	<ol style="list-style-type: none"> <li>1. Pearson correlation coefficient or (<b>PROC CORR</b>)</li> <li>2. Spearman's correlation coefficient (<b>PROC CORR</b>)</li> </ol>
Survival data	<ol style="list-style-type: none"> <li>1. Kaplan-Meier survival curve (<b>PROC LIFETEST</b>)</li> <li>2. Log-rank test (<b>PROC LIFETEST</b>)</li> <li>3. Cox's proportional hazard model (<b>PROC PHREG</b>)</li> </ol>
Agreement Study	<ol style="list-style-type: none"> <li>1. Sensitivity, Specificity (<b>PROC FREQ</b>)</li> <li>2. Kappa measure (<b>PROC FREQ</b>)</li> <li>3. Concordant correlation coefficient</li> <li>4. ROC Analysis including AUROC curve</li> <li>5. Bland-Altman analysis</li> </ol>

# 코호트 연구 (Cohort Study)

- 코호트 연구란 특정 병인학적 요인(etiological factor)에 노출되는 것이 질병발생에 영향을 미치는지의 여부를 파악하기 위한 목적으로, 일련의 연구 대상자 집단을 시간이 흐름에 따라 조사해 나가는 연구
- 처리를 할당하지 않는다. 즉, 관찰연구.
- 편향이 발생할 수 있다.
- 일반적으로 코호트 연구는 전향적이나 후향적 연구도 가능하다. 이 경우 기억이나 자료에 근거하므로 편향이 발생할 가능성이 높다.
- 고정적 코호트(fixed cohort): 연구 대상자가 떠나거나 추적하지 못하는 경우 새로운 연구대상자로 대체하지 않는다.
- 역동적 코호트 (dynamic cohort): 연구 대상자가 떠나거나 추적하지 못하는 경우 자격조건을 갖춘 새로운 연구 대상자로 대체한다.

# 코호트 연구 (Cohort Study)



# 코호트 연구 (Cohort Study)

- 코호트의 선택

- 대상 인구집단의 대표성
- 연구 시작점에서 관심 질병과는 무관한 사람들로 구성
- 건강한 사람들로 이루어진 코호트는 종종 건강인 진입효과 (healthy entrant effect, 또는 healthy worker effect)를 보인다.
- 건강인 진입효과: 연구 시작점에서 연구대상자들이 보이는 사망률이 일반 인구집단에서 기대되는 사망률보다 더 낮아지게 되는 현상

# 코호트 연구 (Cohort Study)

- 연구대상자들의 추적 (follow-up)
  - 이사, 참여거부 등 여러 가지 이유로 추적실패 (lost to follow-up)가 되는 경우가 많다.
  - 연구기간이 길수록 추적실패는 많아진다.
  - 이런 추적실패는 연구의 질을 낮추고 편향을 발생시키므로 연구대상자의 탈락 (drop-out)을 막기 위한 많은 노력을 기울여야 한다.
- 연구진행시 반복적으로 연구의 관심요인들을 조사한다.

# 코호트 연구 (Cohort Study)

- 장점

- 시간에 따른 사건의 진행사항을 알 수 있다.
- 다양한 정보를 확보할 수 있다.
- 질병의 발생률이나 위험도를 직접 추정할 수 있다.
- 희귀한 요인에 노출되는 연구가 가능
- 시간에 따른 노출의 변화에 관한 연구 가능
- 사례-대조 연구에 비해 기억편향(recall bias)이나 선택편향(selection bias) 등이 준다.

- 단점

- 비용과 노력이 많이 든다.
- 희귀질병이 관심인 경우 매우 큰 표본수가 필요
- 추적실패 또는 탈락의 증가
- 장기간 연구로 측정의 일관성 유지 어려움
- 질병의 결과나 병인이 시간에 따라 변화할 수 있다.

# 사례-대조 연구 (Case-Control Study)

- 특정 질병을 가지고 있는 환자집단 (사례군: cases)과 이 질병이 없는 집단 (대조군: controls)의 특성을 비교하여 질병발생 위험을 증가시키거나 감소시키는 요인에 대한 단서를 파악
- 후향적(retrospective) 연구 – 유병률이나 발생률에 관한 정보는 파악불가
- 코호트 연구보다 비용이 적게 든다.
- 희귀질환의 연구에 자주 쓰인다.
- 병과 병인의 관계가 잘 알려지지 않은 경우 예비연구(preliminary study)로도 쓰인다.



# 사례-대조 연구 (Case-Control Study)

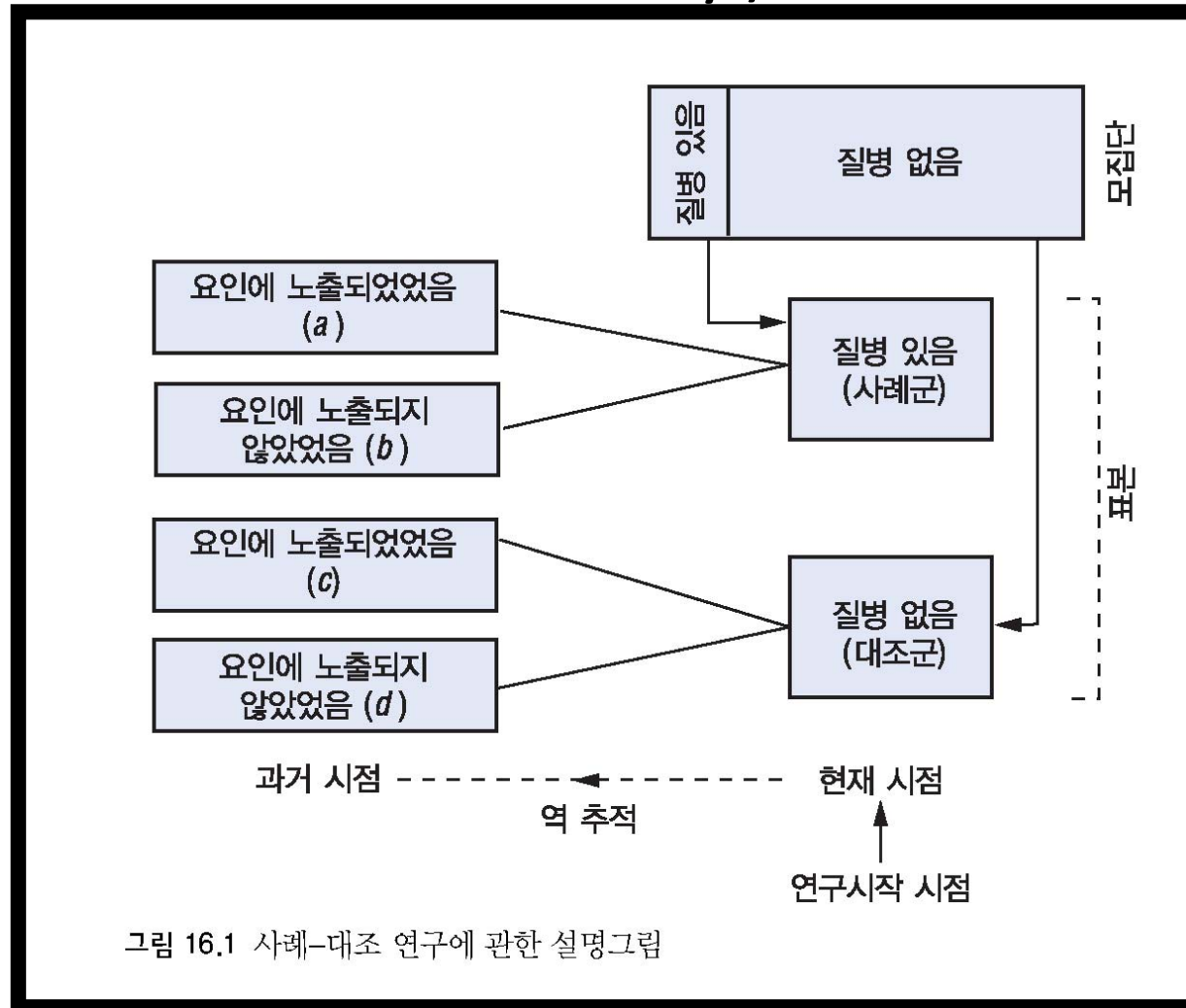


그림 16.1 사례-대조 연구에 관한 설명그림

# 사례-대조 연구 (Case-Control Study)

- 사례군의 선택
  - 발생건(incident case) 또는 유병건(prevalent case) 구분
  - 많은 수의 사례군을 확보하는 것이 중요
- 대조군의 선택
  - 선별검사 (관심 질병이 없음을 확인)
  - 관심 질병 이외의 다른 요인들이 사례군과 비슷한 집단에서 선택

# 사례-대조 연구 (Case-Control Study)

- 짝짓기 (matching)
  - 사례군과 대조군이 유사한 특성을 가지게 하기 위해 사용
    - 연구설계단계에서 교란효과의 통제
  - 연구관심 요인 또는 질병의 병인과 관련된 요인을 대상으로 짝짓기를 하지 않는다.
  - 종류
    - 개인별 짝짓기 또는 1:1 짝짓기
    - 빈도 짝짓기
  - 장점
    - 교란효과의 직접통계
    - Matching이 추정의 정밀도를 높일 수도 있다.
  - 단점
    - 자료의 수집이 복잡해진다.
    - 통계분석은 이 matching을 고려해야 한다.
    - Matching 변수의 효과는 알 수 없다.

# 사례-대조 연구 (Case-Control Study)

- 짝지은 사례-대조 연구
  - 짝지은 변수를 기반으로 사례군과 대조군이 상관되어 있다.
  - 분석에서 이 상관을 고려해 주어야 한다.
  - McNemar's test, Conditional logistic regression 등 특별한 통계기법이 사용되어야 한다.

# 사례-대조 연구 (Case-Control Study)

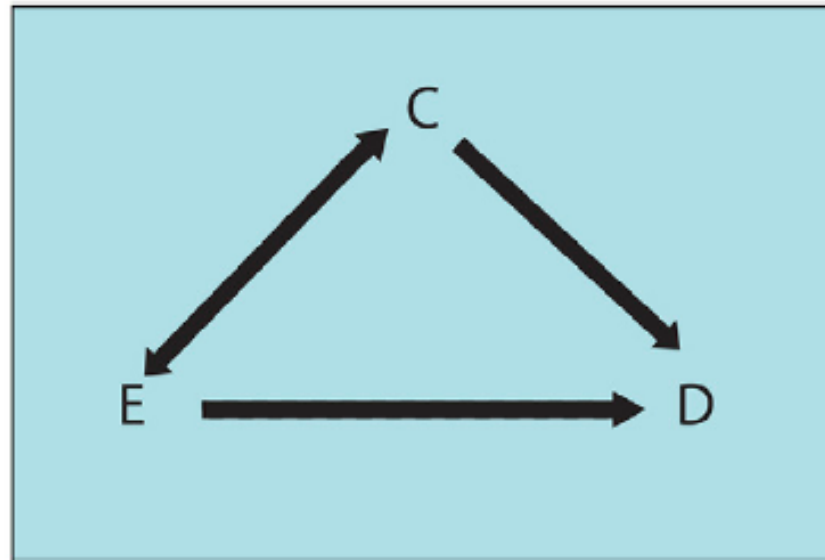
- 장점

- 일반적으로 빠르고 비용이 적게 든다.
- 희귀질병의 연구에 효과적
- 다양한 위험요인들을 조사할 수 있다.

- 단점

- 기억편향(recall bias): 아픈 사람이 병인을 더 잘 기억
- 인과관계의 추론이 어렵다.
- 노출이 희귀한 경우 사용하기 어렵다.

# 교란편향 (Confounding Bias)



- 교란변수(C: confounder)에 의한 처치(E: Exposure)와 병(D: Disease)의 관계 도식화

# 교란편향을 보정하는 통계적 방법

- 짝짓기 (Matching)

- 처치군과 대조군에서 교란변수의 분포를 동일하게 하는 개체를 선택하는 방법 (예: one-to-one matching)

- 장점

- 교란변수의 각 수준에서 처치군과 대조군의 수가 적절하게 보장된다면 이 방법은 교란편향을 최소화할 수 있는 효율적인 방법
- Matching 후의 분석방법은 비교적 쉬움. (처치 여부를 매칭하는 경우)

- 단점

- 교란변수의 수가 증가할 수록 짝지을 개체를 찾기가 어려움
- 짝지은 자료만으로도 충분한 검정력을 갖도록 표본의 수가 커야함.
- 측정되지 않은 교란변수의 통제는 불가능
- 사례-대조 연구의 경우 Over-matching의 문제가 있을 수 있음.
- 사례-대조 연구의 매칭의 경우 매칭의 영향을 분석에서 고려해야 함 (예: McNemar's test, conditional logistic regression)

# 교란편향을 보정하는 통계적 방법

- 층화 (Stratification)
  - 교란변수의 균일 층(homogeneous strata)내에서 처치와 결과를 평가함.
  - 예제

**TABLE II Stratified Analysis of Nail Fixation Compared with Plate Fixation and the Effect on Development of Adult Respiratory Distress Syndrome or Multiple Organ Failure<sup>17</sup>**

	Chest Injury		No Chest Injury		Total Cohort	
	Nail	Plate	Nail	Plate	Nail	Plate
Developed adult respiratory distress syndrome or multiple organ failure	5	2	4	1	9	3
Number at risk	117	104	118	114	235	218
Risk	0.043	0.019	0.033	0.009	0.038	0.014
Risk difference (95% confidence interval)	0.024 (-0.022, 0.069)*		0.025 (-0.012, 0.062)*		0.024 (-0.004, 0.054)	
Summary risk difference (95% confidence interval)	0.024 (-0.005, 0.053)†				P value = 0.11‡	

\*Risk differences between strata are not significantly different; that is, no interaction (test for heterogeneity; p value = 0.96). †Given the absence of interaction, a pooled summary validly estimates the risk difference, adjusting for chest injury. ‡P-value testing the null hypothesis of no association between treatment method and outcome, adjusting for chest injury and assuming no interaction.

Saam Morshed et al. Analysis of Observational Studies: A Guide to Understanding Statistical Methods. J Bone Joint Surg Am. 2009;91 Suppl 3:50-60



# 교란편향을 보정하는 통계적 방법

- 층화
  - 장점
    - 간편함
    - Effect modification을 볼 수 있음
  - 단점
    - 여러 개의 교란변수를 통제하기 어려움
    - 연속형 교란변수의 통제가 어려움
- 멘텔-헨젤 (Mantel-Hanszel Statistics)

# 교란편향을 보정하는 통계적 방법

- 다변수 회귀분석 (Multivariable Regression)
  - 여러 개의 독립변수와 결과변수와의 함수적 관계를 모형화하여 교란변수를 통제함.
  - 예제: 결과변수의 형태에 따른 회귀모형

결과변수의 형태	회귀분석 모형	효과의 추정치
연속형	선형 회귀	평균의 차이, 기울기
이분형	로지스틱 회귀	오즈비 (odds ratio)
사건까지의 시간	콕스의 비례위험	위험비 (hazard ratio)
율 (rate)	포아송 회귀	율의 비 (rate ratio)

# 교란편향을 보정하는 통계적 방법

- 다변수 회귀분석

- 장점

- 여러 개의 교란변수를 효율적으로 통제할 수 있음.
    - 개개 인자의 효과를 비교적 쉽게 평가할 수 있음.

- 단점

- 모형의 선택에 따라 결과가 민감할 수 있음.
    - 잘못된 모형 선택 시 편향이 더욱 심해질 수 있음.
    - 모형의 가정에 민감할 수 있음
    - 다중공선성 등 수학적 문제 발생 가능성

# 교란편향을 보정하는 통계적 방법

- 성향점수 짝짓기 (Propensity Score Matching: PSM)
  - 처치군에 속할 확률에 따라 짝짓기
  - 짝지을 변수가 너무 많은 경우 일반 짝짓기 방법은 현실적으로 불가능
  - 성향점수에 근거하여 짝짓기
    - 성향점수 =  $\text{Pr}(\text{Treatment} \mid \text{Confounders})$
  - $0 \leq \text{성향점수} \leq 1$
  - 성향점수 이론에 의하면 성향점수가 주어진 경우에 편향이 없다면 교란변수가 주어진 경우에도 편향이 없음.

# 교란편향을 보정하는 통계적 방법

- 성향점수 짝짓기 (Propensity Score Matching: PSM)
  - 성향점수는 알려져 있지 않기 때문에 자료로부터 예측하여야 함.
  - 처치가 2개의 수준으로 구성된 경우에는 일반적으로 로지스틱 회귀분석이나 프로빗 회귀분석을 이용하여 성향점수를 예측함.

$$\ln\left(\frac{PS}{1-PS}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$
$$PS = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}$$

# 교란편향을 보정하는 통계적 방법

- PSM의 절차

- 성향점수 모형을 구축함.
- 모든 자료를 이용하여 성향점수를 예측함.
- 짝짓기 알고리즘과 성향점수를 이용하여 처치군과 대조군을 짝지음.
- 짝지은 자료에서 교란변수들이 군간 균형을 이루는 지 평가함. 만일 균형이 충분하지 않은 경우 성향점수 모형을 재구축함.
- 짝지은 자료를 이용하여 평균처치효과를 추정함.
- 이 이외에도 층화, 성향점수 회귀보정, 성향점수 가중치 방법도 있음.

# 교란편향을 보정하는 통계적 방법 예제

- 결혼여부와 우울증과의 관련성
- 대상: 2009년 **지역사회건강조사 자료**를 바탕으로 전국에 살고 있는 만 25세 이상의 남자, 여자 172,828명
- 우울증은 의사진단여부에 따라 우울증 여부를 판단한 변수이고, 결혼여부는 대상자에서 이혼과 사별을 제외한 기혼과 미혼.
  - 주 결과 변수: 우울증 여부
  - 중앙 추적기간: 3.1년
  - 단변수 분석(unadjusted analysis): Unadjusted odds ratio – 0.73 (0.65, 0.81)
  - 미혼이 기혼보다 우울증에 걸릴 오즈(위험)는 약 27% 감소한다. (?)

# 교란편향을 보정하는 통계적 방법 예제

- 잠재적 교란변수
  - 나이, 성별, BMI, 가구원수, 가구소득, 평균수면시간, 지역 등 13개 변수
  - 보정전 교란변수 비교

변수	기혼	미혼	p-value**
나이	50.79±13.46	32.42±8.02	<.0001
BMI	23.16±2.93	22.42±3.19	<.0001
가구원수	3.31±1.3	2.96±1.42	<.0001
가구소득	278.7±280.2	280.2±269.9	0.4556
평균수면시간	6.75±1.25	6.82±1.19	<.0001
성별			<.0001
	남자	75111(49.74%)	13937(63.86%)
	여자	75891(50.26%)	7889(36.14%)



# 교란편향을 보정하는 통계적 방법 예제

- 로지스틱 회귀분석으로 보정
  - 모든 공변량을 회귀모형에 넣어 보정
  - Odds ratio가 1.47 (95% CI: 1.30-1.67)로 미혼이 기혼보다 우울증에 걸릴 Odds가 1.47배 높다.

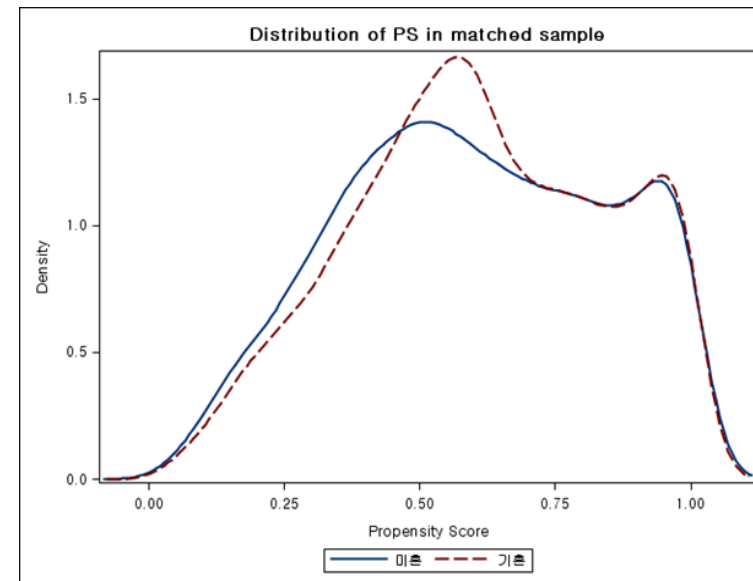
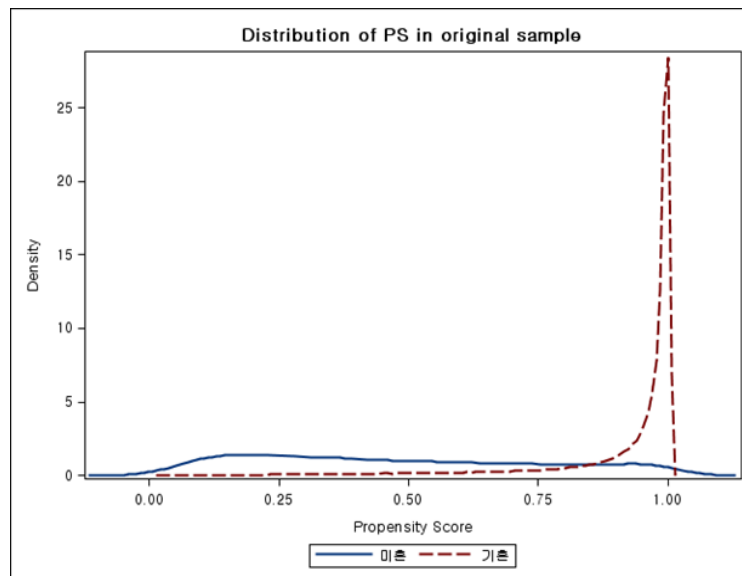
# 교란편향을 보정하는 통계적 방법 예제

- 성향점수 짝짓기 방법

- 성향점수는 13개 잠재적 교란변수를 독립변수로 결혼여부를 종속변수로 하여 로지스틱 회귀모형으로 예측
- Greedy 짝짓기를 하기 위해 로짓의 표준편차에 0.2를 곱한 캘리퍼의 폭은 5.27을 구하여 14,994쌍을 얻어 표본수가 29,988명으로 자료의 약 17.35%를 사용하여 분석을 실시
- Odds ratio가 1.56 (95% CI: 1.30-1.85)로 미혼이 기혼보다 우울증에 걸릴 Odds가 1.47배 높다.
- 균형성 검토 결과 표준화된 차이가 0.1미만이면 균형성을 만족한다고 할 수 있음. 이 분석에서는 대부분 표준화된 차이가 0.1미만으로 균형성을 만족

# 교란편향을 보정하는 통계적 방법 예제

- Distribution of the propensity score in treated and untreated subjects before and after matching



# 교란편향을 보정하는 통계적 방법 예제

성향점수 매칭 후 교란변수 비교

Variable	기혼*	미혼*	p-value**
나이	34.61±7.71	34.31±8.81	<b>0.0018</b>
BMI	22.57±3.04	22.55±3.3	0.5778
가구원수	3.08±1.14	3.13±1.48	<b>0.0006</b>
가구소득	285.9±179.9	290.7±303.9	0.0955
평균수면시간	6.85±1.17	6.85±1.21	0.8043
성별			0.5987
남자	8652(57.7%)	8697(58%)	
여자	6342(42.3%)	6297(42%)	
도시			0.9705

# 교란편향을 보정하는 통계적 방법 예제

- 보정방법에 따른 오즈비 추정치

Odds Ratio Estimates				
Method	Odds Ratio	95% C.I. for OR		p-value
Unadjusted	<b>0.72</b>	0.65	0.81	<.0001
Regression Adjusted	<b>1.47</b>	1.30	1.67	<.0001
PS stratified	<b>1.45</b>	1.28	1.64	<.0001
PS Regression (Categorical)	<b>1.50</b>	1.31	1.73	<.0001
PS Regression (Continuous)	<b>1.74</b>	1.52	1.99	<.0001
Matched	<b>1.56</b>	1.30	1.85	<.0001

# 교란편향을 보정하는 통계적 방법

- PSM 예제: 결혼여부와 우울증과의 관련성
- 결론 및 제약점
  - 이 연구의 표본은 아마도 RCT보다 좀 더 실제 모집단을 대표할 수 있는 표본임.
  - PSM은 RCT가 아님
  - 단지 측정된 교란변수만 보정이 가능함.
    - 민감도 분석
  - 처치와 대조군 사이에 겹치는 부분이 많아야 함.
  - 큰 표본수가 필요함.
- 성향점수를 이용한 다른 방법
  - 가중치 방법
  - 회귀 보정

# 결론

- 관찰연구에서는 선택편향 또는 교란편향을 최소화할 수 있는 방법을 선택해야 한다.
- 각 방법의 장, 단점을 잘 이해하고 가장 적절한 방법을 선택하여 분석을 실시한다.
- 방법을 적용하는 경우 방법의 가정을 만족하는 지 꼭 확인하여야 한다.
- 결과에서 인과관계(causality)를 해석할 때는 연구설계, 편향의 최소화 등 다각도로 고려해야 하며 관찰연구에서의 제한점을 꼭 숙지하고 있어야 한다.
- 관찰연구가 종적(longitudinal)이면 중도탈락 등으로 인한 결측이 많이 발생할 수 있다. 결측은 또다른 편향을 발생시킬 수 있으므로 결측을 고려한 분석법을 적용하여야 한다.