

임상시험 통계의 이해

울산의대 서울아산병원

의학통계학교실

윤성철

임상연구의 정의

- ◆ 직접 사람(환자)를 대상으로 하는 의학연구
 - 질병의 발생에 영향을 미치는 위험요인 규명
 - 진단법의 타당도와 신뢰도 평가
 - 치료법/약물의 효능과 안전성 평가
 - 질병의 자연사 연구

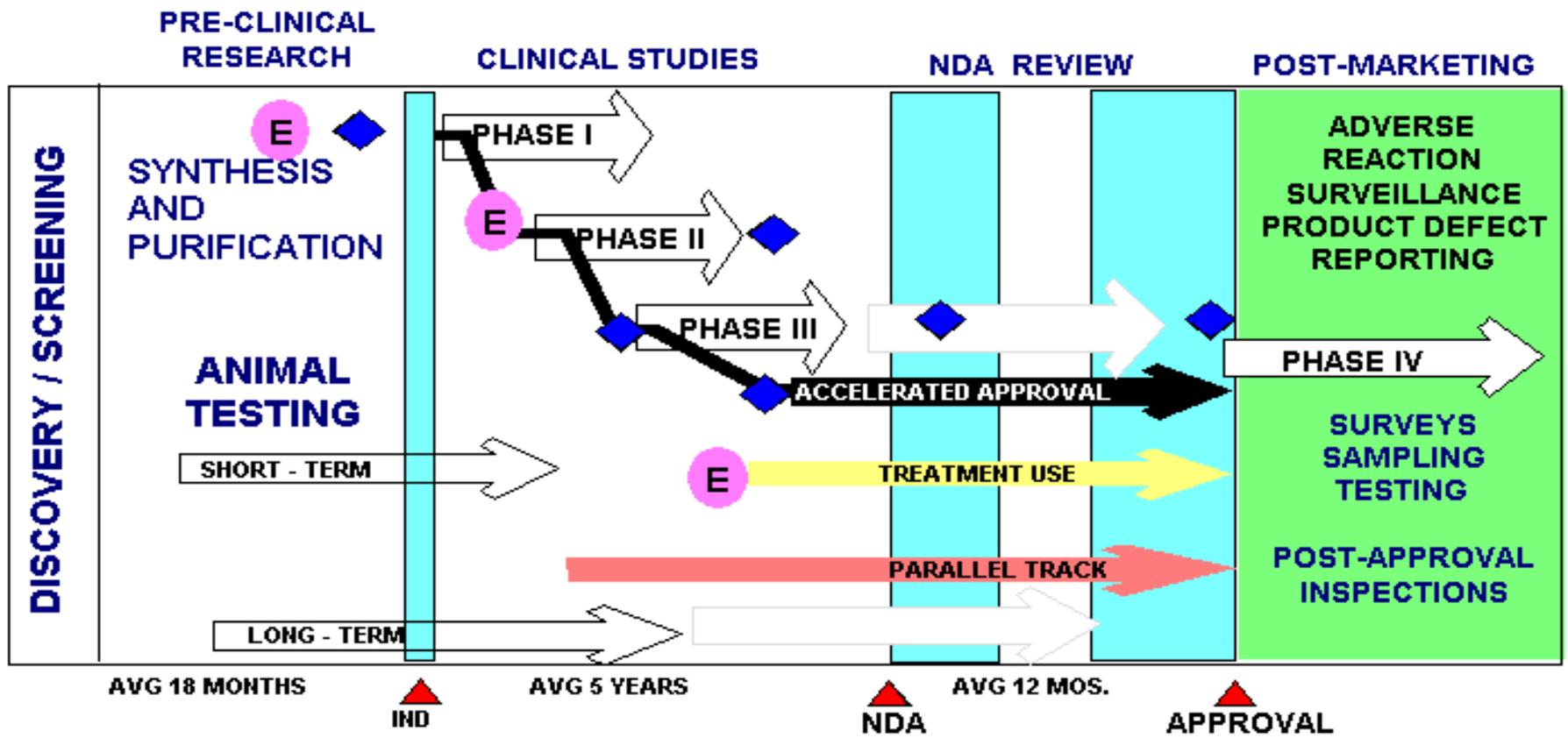
임상연구 설계의 종류

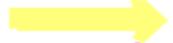
연구설계의종류	연구방법	설득력크기	수행빈도
환자사례보고 (Case report)	관찰	가장 약함	가장 빈번
환자군연구 (Case series study)	관찰		
단면적연구 (cross-sectional study)	관찰		
환자-대조군연구 (case-control study)	관찰		
코호트연구 (cohort study)	관찰		
무작위배정 임상시험 (Randomized clinical trial)	실험	가장 강함	가장 드뭄

임상시험에서의 통계의 역할

- ◆ 임상시험은 **의약품/의료기기**의 **안전성과 유효성**을 평가하기 위한 구조화된 연구형태
- ◆ 임상시험에서 통계학은 다음 사항을 제공
 - 타당한 결과를 얻을 수 있도록 **연구설계**
 - 안전성과 유효성을 **정량화**할 수 있도록 구조화
 - 다른 의약품과 **비교**할 수 있도록 구조화
 - **신뢰성**있고 **재생가능**한 결과 도출

Drug Development Process



	FDA TIME	ACCELERATED REVIEW:	EXPANDED ACCESS:
	INDUSTRY TIME	 SUBPART E	 TREATMENT USE
	FDA & INDUSTRY TIME	 ACCELERATED APPROVAL	 PARALLEL TRACK
	SPONSOR/FDA MEETINGS ENCOURAGED		

Courtesy of Jang IJ

임상시험의 Quality

◆ 과학성

- 목적의 필요성
- 방법의 타당성 : 치료군 간의 비교성 / 결과의 일반화 가능성
 - 결과변수의 선택과 측정 : Primary Outcome
 - 항고혈압제 : 혈압 ; 골다공증 : 골절; 항암제 : 5YSR, DFS 등
- 통계적 건전성 : 표본크기 및 분석방법의 적절성

◆ 윤리성

- 동의서, 설명문, 사회적 가치, 피험자의 존중 등..

◆ 신뢰성

- 임상시험관리기준(GCP) 준수 : 문서화, 표준작업지침서(SOP)
- 좋은 계획서/증례기록서(CRF)
- 질적 자료관리 (CDM)
- 품질보증 (QA) : Monitoring, Audit, Inspection

임상시험에서의 통계사용 영역

◆ 임상시험 연구설계 단계

- 1차/2차 결과변수 설정 : Validity / Reliability
 - 결과변수가 진단, 치료, 예후를 가장 잘 반영하는가
- Design 설정 : Parallel, Crossover 등...
 - 우월성, 동등성, 비열등성, 차이성 (inequality)
 - 비교성 : Randomization 및 blinding 등
- 표본 크기 및 **분석 대상군**의 정의 및 **결측처리**

◆ 임상시험 자료분석 단계

- 임상시험은 전향적 실험연구이므로 상대적으로 **통계분석방법이 단순**

임상시험에서 통계분석 일반원칙

- ◆ 연구계획서상에 **통계분석계획에 기술된 대로** 분석하는 것을 원칙으로 한다.
- ◆ **통계분석계획(SAP)**
 - 주결과변수, 이차결과변수에 대한 분석방법
 - 분석대상군 : ITT, FAS, PP data set
 - 분석상의 문제점 해결방안
- ◆ **눈가림 해제 이전에 검토가능**: 문서화 되어있어야 함.
- ◆ 임상시험계획서에 기술된 통계분석계획으로부터 얻어진 결과만이 해당 임상시험 계획서에 대한 결과로써 인정받을 수 있다.

임상시험 통계분석 예제

Table 16-6 Differences in pretherapy and posttherapy heart rate, in beats per minute (HR), for groups *A* and *B*, with 25 participants each

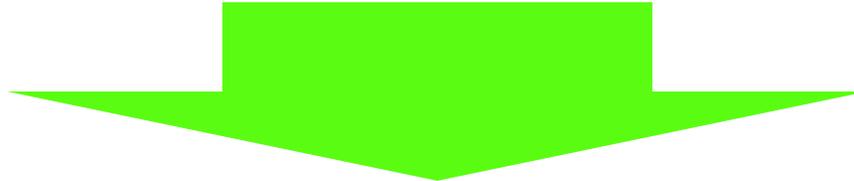
Observation number	<i>A</i>			<i>B</i>		
	Pre-HR	Post-HR	Change in HR	Pre-HR	Post-HR	Change in HR
1	72	72	0	72	70	2
2	74	73	1	71	68	-3
3	77	71	6	75	74	1
4	73	78	-5	74	71	3
5	70	66	4	71	73	-2
6	72	76	-4	73	78	-5
7	72	72	0	71	69	2
8	78	76	2	70	74	-4
9	72	80	-8	79	78	1
10	78	71	7	71	72	-1
11	76	70	6	78	79	-1
12	73	77	-4	72	75	-3
13	77	75	2	73	72	1
14	73	79	-6	72	69	3
15	76	76	0	77	74	3
16	74	76	-2	79	75	4
17	71	69	2	77	75	2
18	72	71	1	75	75	0
19	68	72	-4	71	70	1
20	78	75	3	78	74	4
21	76	76	0	75	80	-5
22	70	63	7	71	72	-1
23	76	70	6	77	77	0
24	78	73	5	79	76	3
25	73	73	0	79	79	0
Mean	73.96	73.2	0.76	74.40	73.96	0.44
Standard deviation	2.88	3.96	4.24	3.18	3.38	2.66

: A약과 B약에서
치료 이후 맥박수 측정

 P=0.7512

Use of cutpoints

Beats/min	< 7	≥ 7	< 5	≥ 5	< 3	≥ 3
Group A	25	2	19	6	17	8
Group B	25	0	25	0	18	7
Chi-square	$p = 0.15$		$p = 0.009$		$p = 0.76$	
Fisher's exact	$p = 0.49$		$p = 0.022$		$p = 0.99$	



분석계획이 사전에 정의되어 있어야 한다.

연구가설

통계적 추론 : 가설 검정

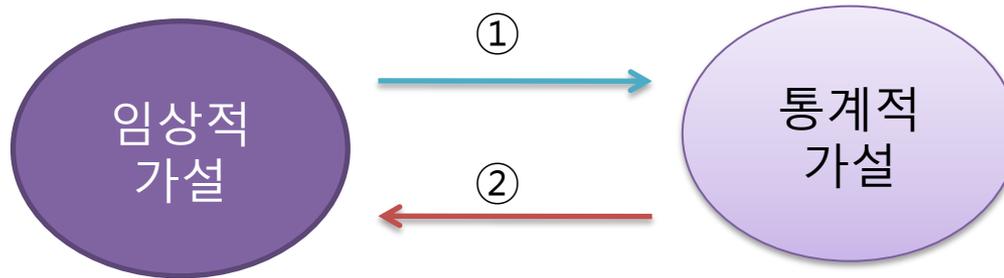
- ◆ **가설검정** : 표본자료를 이용하여 통계적인 방법으로 모집단의 특성에 대한 주장을 기각하거나 채택하는 의사결정을 말한다.
- ◆ **대립가설** : 표본을 통하여 얻어지는 정보의 강력한 시사에 힘입어 모집단에 대한 예상이나 주장. (\leftrightarrow 귀무가설)

사실	가설검정 결과	
	H_0 채택	H_0 기각
H_0 참	옳음	제1종 오류(α)
H_A 참	제2종 오류(β)	옳음

◆ 양측검정과 단측검정

- 양측검정(two-sided test): 대립가설이 '... 다르다'고 설정
- 단측검정(one-sided test): 대립가설이 '... 보다 크다' 또는 '... 보다 작다'

임상적 가설 VS. 통계적 가설



예) A 약과 B 약의 효과의 차이가 있는가?

- Overall survival 의 차이?
- 1 year survival, 5 year survival 등

예) 치료방법간에 성공률이 다른가?

- $P_a = 44\%$ vs. $P_b = 42\%$: Diff. = 2% RR= 1.05
- $P_a = 4\%$ vs. $P_b = 2\%$: Diff. = 2% RR= 2

임상시험에서의 가설

◆ 우위성 평가시험(superiority trials)

- $H_0: p_t \leq p_c + \Delta$ vs. $H_1: p_t > p_c + \Delta$
- Δ (margin of equivalence, >0)

◆ 비 열등성 평가시험(non-inferiority trials)

- $H_0: p_t \leq p_c - \Delta$ vs. $H_1: p_t > p_c - \Delta$

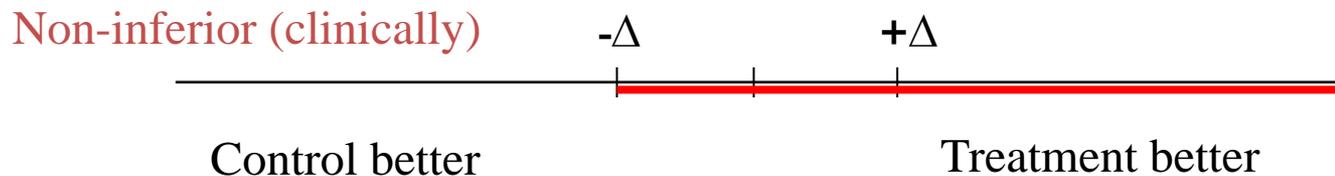
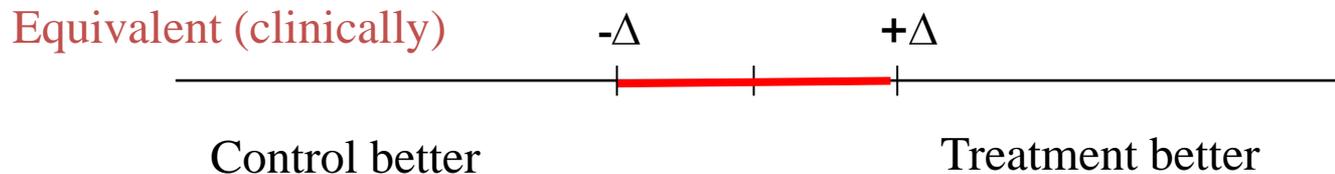
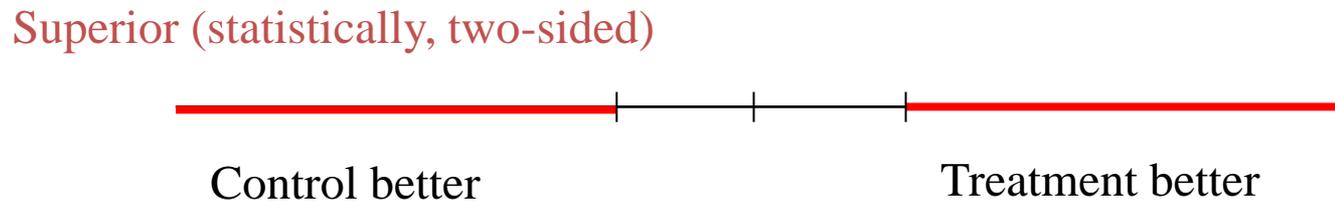
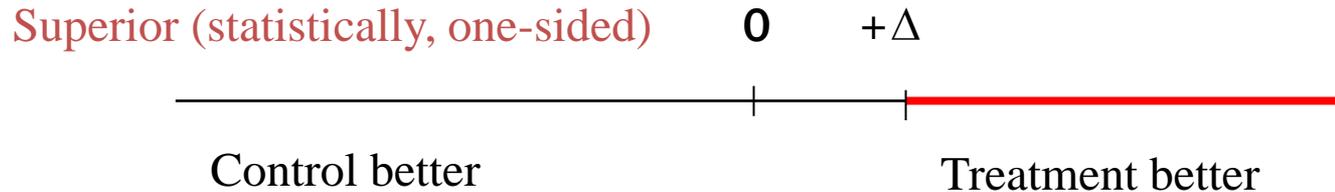
◆ 동등성 평가시험(equivalence trials)

- $H_0: |p_t - p_c| \geq \Delta$ vs. $H_1: |p_t - p_c| < \Delta$
- 타 검정에 비해 더 많은 대상수가 필요

Choice of Δ

- ◆ A different choice of Δ may
 - affect the sample size calculation
 - alter the conclusion
- ◆ 우월성
 - a standard effect size between 0.25~0.5 usually chosen
- ◆ 비열등성 : binary responses
 - 50~80% $\rightarrow \Delta=20\%$, 80~90% $\rightarrow \Delta=15\%$
 - 90~95% $\rightarrow \Delta=10\%$, >95% $\rightarrow \Delta=5\%$
- ◆ Bioequivalence trials with healthy volunteers
 - $\Delta = \log(1.25)$ for mean difference on log-transformed data

Superior/Equivalent/ Non-inferior Test



One-Sided vs. Two Sided Test

- ◆ Sample size may be reduced by about 15~25%
- ◆ In practice a two-sided test may be preferred
 - because placebo effect may be substantial in many drug products such as drug products regarding diseases in the central nervous system
- ◆ Doby(1991) : one-sided test procedure may be justified
 - (i) Toxicity studies (ii) safety evaluation
 - (iii) Analysis of ADR data (iv) risk evaluation
 - (v) Laboratory data

피험자 수

중요성 및 필요성

- ◆ 피험자 수가 부족하면
 - 실제 존재하는 중요한 효과를 찾아내지 못한다.
 - = 통계적 **검정력**이 떨어지므로 의약품에 효과가 있는데도 이를 찾아내지 못한다.
- ◆ 피험자 수가 불필요하게 많으면
 - 시간과 돈의 낭비이다.
 - 안전성 및 유효성이 입증되지 않은 의약품이라면 **윤리적으로 문제가 있다.**
- ◆ 따라서 ...
 - 피험자 수에 영향을 주는 요인(유의수준, 검정력, 치료 효과의 크기와 변이 등)을 감안하여 피험자 수를 최적화하여야 한다.
 - 특히 '치료효과(유효성)의 크기와 변이', 즉 임상시험 결과에 대한 사전 예측이 필수적이다

피험자 수 계산의 원칙과 고려 사항

- ICH guideline E9 -

- ◆ 임상시험을 통하여 해결하고자 하는 질문에 신뢰성 있는 답을 제공할 수 있도록 충분히 커야 함
- ◆ 일반적으로 임상시험의 **일차 목적(primary outcome)**에 근거하여 결정
- ◆ **안전성**이나 중요한 **이차 목적**에 근거할 수도 있지만 이 경우 그 이유는 명확하고 정당하여야 함

피험자 수 계산의 원칙과 고려 사항

- ICH guideline E9 -

◆ 피험자 수 결정에 고려되는 일반적인 항목

- 일차 변수의 형태
- 통계적 가설 (귀무가설 vs. 대립가설)
- 검정 통계량 : 실험설계 방법, 일차변수 형태에 의존
- 제 1종 오류와 제 2종의 오류
- 탈락과 계획서 위반을 처리하는 접근법

◆ 계획서의 피험자 수 계산에 기술되어야 내용

- 피험자 수 계산에 사용되는 추정치들
 - 평균과 분산 또는 반응률, 치료효과 차이의 크기
 - 발표된 자료나 과거 임상시험에 근거
- 추정치의 변화에 따른 피험자 수 크기변화 평가
- 제 1종 오류와 제 2종의 오류 크기
 - 통상적으로 제1종 오류 5% 이하, 제2종 오류 10% 또는 20%

피험자 수 계산

◆ 피험자 탈락률의 반영

- 앞의 계산을 통하여 구한 피험자 수가 n 이고 예상되는 임상시험 기간 동안 피험자 탈락률이 a 라고 하면 임상시험에서 요구되는 실제 피험자 수는 다음과 같이 구한다

$$N = \frac{n}{1 - a}$$

자료분석 대상군

아래와 같은 상황이 주어졌을 때 어떤 원칙으로 문제점을 해결할 수 있을까?

- 배정된 치료를 끝까지 받지 못하였거나 배정된 치료를 받지 않은 대상
- 행정적인 착오 등으로 바뀐 치료를 받은 대상
- 순응도(compliance)가 낮은 대상
- 연구참여의 적격성(eligibility)에 대한 주요 기준을 만족하지 못한 대상
- 일차변수(primary endpoint)에 대한 자료가 없는 대상

ICH E9 : "Statistical principles of Clinical Trials"

V. DATA ANALYSIS CONSIDERATIONS

5.1 Prespecification of the Analysis

5.2 Analysis Sets

5.2.1 Full Analysis Set

5.2.2 Per Protocol Set

5.2.3 Roles of the Different Analysis Sets

5.3 Missing Values and Outliers

5.4 Data Transformations

5.5 Estimation, Confidence Intervals and Hypothesis Testing

5.6 Adjustment of Significance and Confidence Levels

5.7 Subgroups, Interactions and Covariates

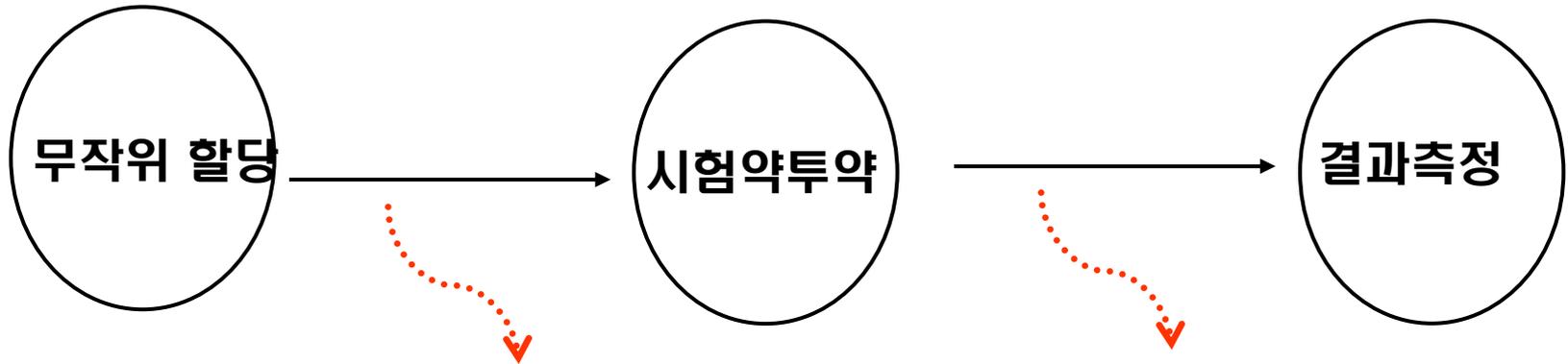
5.8 Integrity of Data and Computer Software Validity

ICH E9 : "Statistical principles of Clinical Trials"

◆ 자료분석 대상군 선택의 원리

- 임상시험계획서에는 다양한 형태의 임상시험계획서 위반, 중도탈락, 결측치 등과 같이 만족할 만한 분석을 실시하는데 영향을 미칠 수 있는 어떠한 불균형의 발생에 대해 이를 최소화하기 위한 절차를 구체적으로 명시하여야 한다.
- 분석대상군은 **비뺄림을 최소화**하고 제 1종 오류가 커지는 것을 억제하는 원칙에 따라 결정되어야 한다
- 분석대상군은 '무작위배정된 피험자군(Full Analysis Set)과 계획서순응 피험자군(Per Protocol Set)이 있음.

통계분석 대상군



subject withdrawal, ineligibility, non-adherence
poor quality data, occurrence of competing event

유효성 분석

- ITT / FAS 분석군 : 배정된 대로 분석
- PP (per protocol) 분석군 : 계획서 순응분석

안전성분석

- ITT (Intention to treat) 분석군 : 안전성분석

(ex) Exclusion & Withdrawl

2.2.1 선정기준

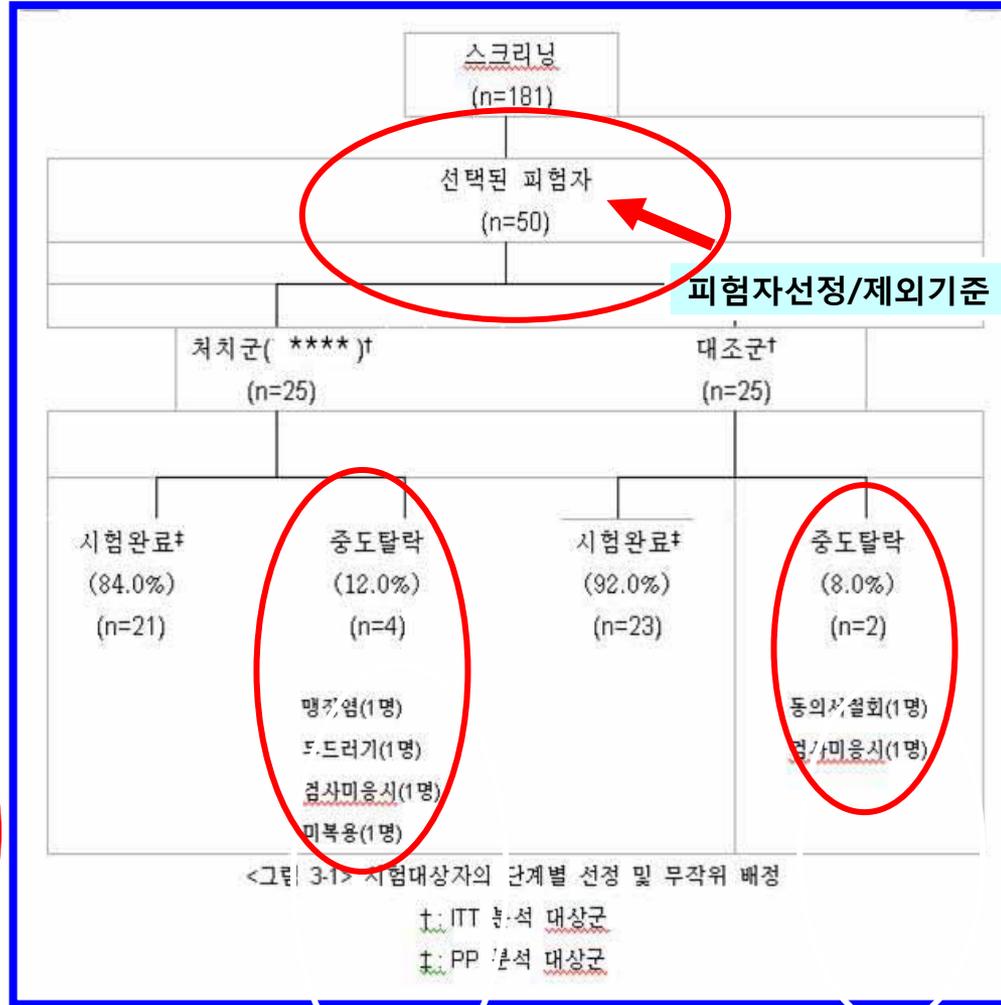
피험자의 선정기준은 성별 구분없이 다음의 기준에 만족하는 경우만 피험자로 선정하였다.

- 1) 만 20세에서 65세 사이인 자.
- 2) 간기능을 나타내주는 수치 가
 $30 \leq \text{AST(SGOT)}$ 또는 $\text{ALT(SGPT)} \leq$ 정상상한치의 2.5배
- 3) 총 Bilirubin이 정상상한치의 1.5배를 넘지 않는 자,
- 4) 연구자로부터 시험의 성격을 충분히 설명받고 자발적으로 서면 동의를 작성한 자.

2.2.2 제외기준

상기 선정기준을 만족하더라도 다음의 제외기준에 하나라도 해당하면 본 인체시험에서 제외하였다.

- 1) 임상적으로 치료를 요하는 간 질환이나 다른 질환을 앓고 있는 자
- 2) 최근 1개월 간 간기능개선제를 복용한 자
- 3) 최근 5년간 악성종양 경험이 있는 자
- 4) 알코올 중독인 자
- 5) 임신 또는 수유 중인 자
- 6) 의약품 또는 기능성 식품을 대상으로 하는 임상시험에 참여한 경험이 있는 자



통계분석 대상군

◆ ITT/FAS 분석군:

- 통계적으로 보수적이다.
- 실제 진료상황에서 관찰할 수 있는 치료효과의 참값에 가까운 값을 구할 수 있다.

◆ PP 분석군:

- 연구계획서를 그대로 준수했을 경우에 관찰할 수 있는 치료효과의 과학적 모형을 잘 반영함.
- 무작위배정상태가 깨지므로 비뚤림 발생가능성

ITT analysis ?

Table 1 Use of intention to treat and other methods to analyse trial of coronary artery bypass surgery and medical treatment for stable angina pectoris in 768 men.² Mortality 2 years after randomisation is shown by allocated and actual intervention*

	Allocated (actual) intervention				Differences in mortality (95%CI) surgical v medical
	Medical (medical)	Medical (surgical)	Surgical (surgical)	Surgical (medical)	
No of survivors	296	48	353	20	—
No of deaths	27	2	15	6	—
Mortality (%)	8.4%	4.0%	4.1%	23.1%	—
Intention to treat analysis	7.8% (29/373)		5.3% (21/394)		2.4% (-1.0% to 6.1%)
Per protocol analysis	8.4% (27/323)		4.1% (15/368)		4.3% (0.7% to 8.2%)
As treated analysis	9.5% (33/349)		4.1% (17/418)		5.4% (1.9% to 9.3%)

*77 patients did not receive allocated intervention for various reasons (one was not available for follow up and is not included in table). The high death rate in the group assigned to surgery but not receiving it is due to 6 patients who died before they could be operated on. The authors correctly reported the intention to treat analysis, which shows no significant difference between the treatments.

BMJ 1999;319:670-4

유효성 평가에 대한 통계분석계획

- ◆ 귀무가설 및 대립가설의 구체적인 기술
- ◆ 주 평가변수 및 이차 평가 변수별 유효성 평가기준 및 방법
- ◆ 통계분석 원칙의 기술
- ◆ 유의수준 및 결과제시와 결론 도출 방안
- ◆ 중간분석 여부 및 구체적인 계획

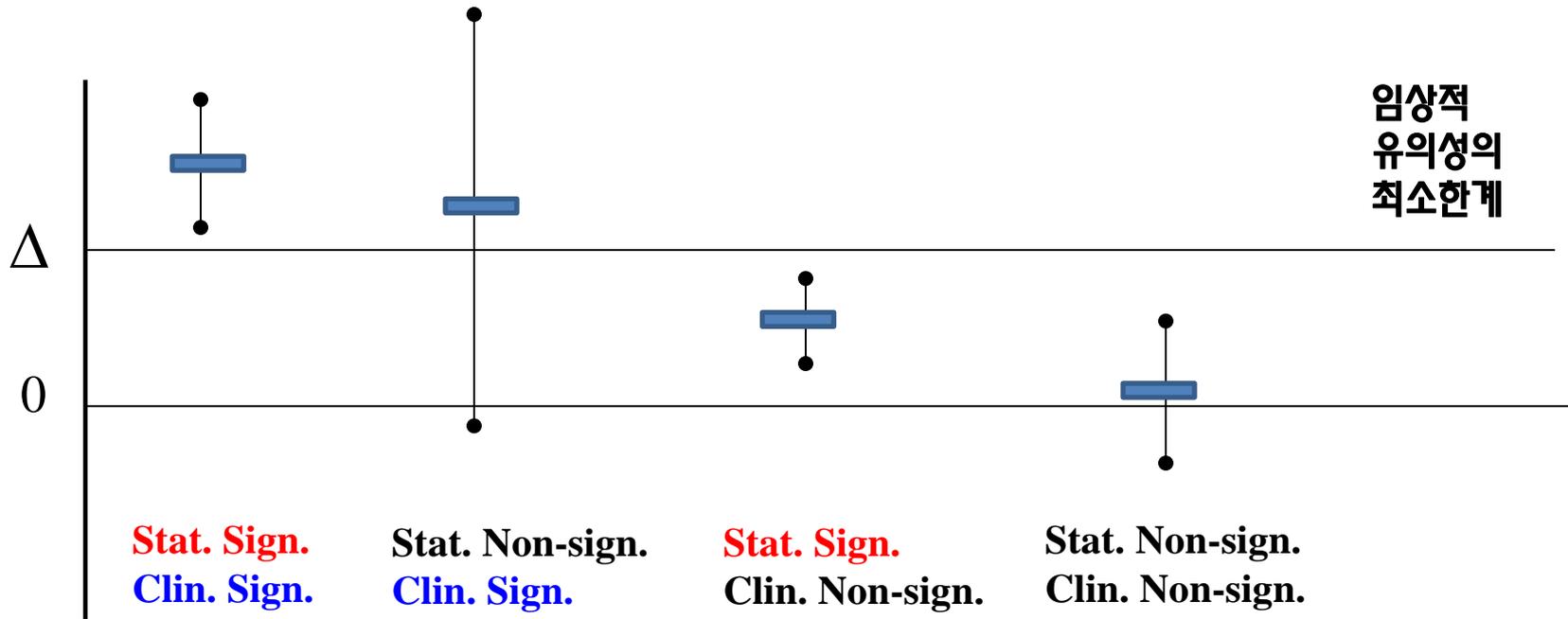
- ◆ 분석대상군의 구체적인 정의
 - ITT분석에서: **결측치 처리방안**에 대한 기술
 - PP분석에서: PP 분석대상군 제시
 - 중도탈락 피험자
 - 순응도

- ◆ 두 군간에 baseline data의 차이 평가방법 및 통계분석시 보정계획

안전성 평가에 대한 통계분석계획

- ◆ 안전성 평가 대상자를 구체화하고 평가 방안을 제시
- ◆ 유해사례 분류 체계를 제시하고 유해사례 발현을 비교 방안을 제시
- ◆ 활력징후 및 실험실적 검사의 변화에 대한 평가 방안 제시

통계적 유의성과 임상적 유의성



임상적 유의성이 통계적 유의성보다 중요하다.

감사합니다...

`ysch97@amc.seoul.kr`