

# 임상연구 설계와 분석을 위한 기본통계

## 차례

서론 .....	7
A. 연구 설계의 기술 .....	9
A-1 연구의 유형 .....	9
A-1.1 연구의 유형: 관찰적 혹은 실험적 연구 .....	9
A-1.2 연구의 결합과 나열 .....	10
A-1.3 코호트 연구 .....	10
A-1.4 환자-대조군 연구 .....	10
A-1.5 단면적 연구 .....	11
A-1.6 교란 .....	13
A-1.7 실험과 시험 .....	13
A-1.8 무작위배정 비교임상시험 .....	14
A-1.9 파일럿연구와 탐색적 연구 .....	14
A-2 추적관찰 .....	15
A-3 연구대상 .....	15
A-4 변수 유형 .....	16
A-4.1 측정의 척도 .....	16
A-4.2 자료의 유형 .....	16
A-4.3 자료 수집의 방법 .....	16
A-4.4 타당도와 신뢰도 .....	17
B. 임상시험 .....	18
B-1 연구대상과 선정제외기준 기술 .....	18
B-2 치료법 기술 .....	19
B-3 대조군의 필요성과 선택방법 .....	19
B-4 눈가림 .....	21
B-4.1 이중눈가림과 단일눈가림 설계 .....	21

B-4.2 위약 .....	22
B-5 무작위배정 .....	22
B-5.1 무작위배정 .....	22
B-5.2 무작위배정이 적용되는 시험 .....	22
B-5.3 무작위배정의 필요성 .....	23
B-5.4 무작위가 아닌 배정방법 .....	23
B-5.5 무작위배정 방법 .....	23
B-5.6 블록 무작위배정 .....	25
B-5.7 층화 무작위배정 .....	25
B-5.8 최소화법 .....	26
B-5.9 군집 .....	30
B-5.10 시험설계 .....	31
B-6 결과변수 .....	32
B-7 데이터 모니터링 .....	33
B-7.1 데이터 모니터링 위원회 .....	33
B-7.2 임상시험 조기종료 시기 .....	34
B-8 동의서 .....	34
B-8.1 동의 .....	34
B-8.2 응급상황 .....	35
B-8.3 어린이 .....	35
B-8.4 스스로 판단할 수 없거나 의사를 표명할 수 없는 피험자 .....	35
B-8.5 집단무작위배정설계 .....	36
B-8.6 무작위배정 후 동의 취득하는 경우 .....	36
B-9 연구계획서 위반과 비순응 .....	37
B-10 연구대상수 달성 .....	37
B-11 임상시험 종료 후 .....	38
<b>C. 관찰적 연구 .....</b>	<b>38</b>
C-1 환자-대조군 연구 .....	38

C-1.1 환자-대조군 연구에서 대조군의 선택 .....	38
C-1.2 환자-대조군 연구에서의 짝짓기 .....	39
C-2 평가 비뚤림 .....	40
C-3 회상 비뚤림 .....	40
C-4 표본조사: 표본 집단의 선택 .....	40
C-5 일반화와 결과의 외삽 .....	41
C-6 설문 조사에서 응답률의 최대화 .....	41
<b>D. 연구대상수 산출 .....</b>	<b>42</b>
D-1 연구대상수 산출이 제시되어야 하는 경우 .....	42
D-2 연구대상수 산출의 중요성 .....	42
D-3 연구대상수 산출에 필요한 정보 .....	42
D-4 통계 용어의 설명 .....	44
D-4.1 귀무가설과 대립가설 .....	44
D-4.2 확률값(p-값, p-value) .....	44
D-4.3 유의수준 .....	44
D-4.4 검정력 .....	44
D-4.5 임상적으로 유의한 효과 크기 .....	44
D-4.6 유의성의 단측검정과 양측검정 .....	45
D-5 연구대상수 산출 시 포함되어야 할 변수 .....	45
D-6 반응률, 탈락률 등에 따른 연구대상수의 손실 .....	45
D-7 연구 목적과 통계적 분석 방법에 따른 연구대상수 산출 .....	45
D-8 연구대상수 산출 및 기술에 대한 예제 .....	46
D-8.1 단일 분율의 추정 .....	46
D-8.2 두 개의 분율 비교 .....	47
D-8.3 두 개의 평균 비교 .....	48
D-9 부적절한 연구대상수 기술 .....	50
<b>E. 통계적 방법의 기술 .....</b>	<b>51</b>
E-1 서론 .....	51

E-1.1 적절한 용어의 사용 .....	51
E-1.2 통계적 방법의 구체적 기술 .....	52
E-2 통계적 방법의 적절성 .....	52
E-2.1 순위형 점수 .....	53
E-3 짝지은 표본의 비교와 독립 표본의 비교 .....	53
E-4 통계적 검정의 가정 .....	54
E-4.1 자료의 변환 .....	54
E-5 교란효과와 보정 .....	55
E-6 계층적(다수준) 자료 .....	55
E-6.1 계층적 자료의 분석 .....	56
E-7 다중검정 .....	56
E-7.1 유의성 검정의 반복적 사용 .....	56
E-7.2 다중검정의 문제점 .....	57
E-7.3 Bonferroni 보정 .....	58
E-7.4 다중검정 문제의 해결 .....	59
E-8 평균값으로의 회귀 현상 .....	61
E-9 임상시험에서의 의도된 대로의 분석원칙 .....	62
E-10 집단 무작위배정 시험 .....	62
E-11 변수의 통합 .....	62
E-12 추정과 신뢰구간 .....	63
E-12.1 분율값이 1 또는 0 에 가까운 경우 .....	64
<b>F. 일반적 고려사항 .....</b>	<b>64</b>
F-1 통계 전문가 (통계 분석) .....	64
F-2 통계 프로그램/패키지 .....	65
F-3 윤리 .....	65
F-3.1 통계의 오용 .....	65
F-3.2 비판적 평가 .....	65
F-3.3 사람을 대상으로 한 연구 .....	65

F-4 기타 연구 쟁점 .....	66
F-4.1 연구의 관리 감독 .....	66
F-4.2 데이터의 보호/기밀유지 .....	66

## 부록

부록 1 임상시험 계획서 작성 시 체크 리스트 .....	67
부록 2 용어 정리 .....	69
참고문헌 .....	74

## 서론

### 배경

의학연구를 수행하기 전에 선행되어야 할 작업이 연구계획서를 작성하는 것이다. 특히 연구계획서 작성은 학위논문심사, 연구비 지원, IRB 심사를 위해 필수적이며 매우 중요한 단계이다. 연구계획서에서는 배경 및 목적, 연구가설, 연구 설계, 연구대상수의 산출근거, 연구가 어떻게 수행될 것인지에 대한 구체적인 방법, 통계분석 계획 등이 기술되어야 한다. 이 중에서 연구대상수의 산출근거 및 통계분석 계획은 의학연구의 이론적 근거와 결과의 일반화를 제시하는 매우 중요한 요소이므로 과학적, 합리적으로 기술되어야 한다. 그러나 연구계획서 작성에 참여하는 많은 사람들 중에 통계전문가가 관여하는 경우는 그리 많지 않으며 연구자들에 대한 통계교육이 부분적으로 실시되고 있으나 보다 많은 사람들이 쉽게 접할 수 있는 방법이 필요하다. 연구계획서를 작성하게 될 많은 연구자들이 손쉽게 읽고 수행할 수 있는 실용적인 지침서를 개발하고 보급하여 의학연구의 질적 수준 향상과 활성화를 도모하고자 하였다.

이 지침서는 영국 런던대학교 세인트조지병원에서 6명의 의학통계학자(Martin Bland (Professor of Medical Statistics), Barbara Butland (Lecturer in Medical Statistics), Janet Peacock (Senior Lecturer in Medical Statistics), Jan Poloniecki (Senior Lecturer in Medical Statistics), Fiona Reid (Lecturer in Medical Statistics), Philip Sedgwick (Lecturer in Medical Statistics))들이 The South East Regional Office (SERO) Research and Knowledge Management Office Directorate의 지원을 받아 발간한 "Statistics Guide for Research Grant Applicants"를 참고로 하여 작성되었다. 이들 6명의 통계학자들은 The South East Research and Development Project Grant Scheme에 대한 연구비 신청 계획서를 일상적으로 검토하는 과정에서 공통적으로 빈번하게 발생하는 통계적 문제점을 접하게 되었으며 계획서를 작성하는 연구자들이 이를 이해하고 극복할 수 있는 방안을 제시하기 위하여 위와 같은 가이드라인을 개발하게 되었다. 현재 서울대학교 의과대학/서울대학교병원 의학연구협력센터에서는 서울대학교병원의 IRB에 제출되는 다양한 연구계획서에 대하여 연구의 설계 및 수행 방법, 통계분석계획 등이 상호 관련성을 잘 반영하여 과학적으로 기술되어 있는지 검토하고 있으며 영국의 통계학자들이 경험한 것과 같이 공통적으로 빈번하게 발생하는 통계적 문제점을 접하고 있다. 그러므로 연구계획서를 검토한 경험을 기반으로 국내 실정에 맞는 임상연구계획서 작성 가이드라인을 개발하고자 하였다.

### 개발과정

Martin Bland 교수로부터 원문 사용과 참고에 대한 협의를 거친 후 의학연구협력센터 통계연구원과 역학연구원 및 예방의학교실 연구원들이 "Statistics Guide for Research

Grant Applicants”의 구성과 내용을 파악하고 매주 회의를 거쳐 수정방안을 논의하고 이를 반영하여 우리 실정에 맞게 보완하는 과정을 거쳐 이 지침서가 완성되었다. 이를 포켓 사이즈 지침서와 웹페이지(<http://mrcc.snuh.org>, <http://mrcc.snu.ac.kr>)로 제작하여 필요한 연구자들이 쉽게 접근할 수 있도록 할 것이다.

## 지침서의 사용

이 지침서는 통계학을 가르치기 위해 설계된 것이 아니라 이미 기본적인 통계학적 지식을 갖춘 사람들에게 추가 정보를 제공하기 위해 설계된 것이다. 예를 들어, 신뢰구간과 유의성 검정에 대한 기본적인 지식은 있으나 통계적 검정력이나 대상수 산출까지는 이해하지 못한다고 가정하였다. 또한, 이 지침서를 읽는 것만으로 계획서를 제출하기 전에 통계학자들과 연구계획에 대해 논의하는 것을 대체할 수 있는 것은 아니다. 통계학자들과 연구계획에 대해 논의하는 것은 강력히 권장되며, 이 지침서는 연구자들이 통계 상담에 가지고 가야 할 적절한 질문과 올바른 정보를 제대로 인식하고, 더불어 통계자문내용에 대한 이해를 도울 수 있을 것이다.

이 지침서를 처음부터 끝까지 읽기보다는 자신의 연구가 포함하고 있지 않는 부분은 무시할 수 있도록 목차와 체크리스트가 작성되었다. 목차나 체크리스트를 잘 활용할 수 있도록 하기 위해 이 책은 각각의 인덱스 코드(예, A-1.1)가 있는 짧고 완비된 문단들로 구성되어 있다. 이러한 코드들이 검토자, 상담하는 통계학자, 연구자들에게 똑같이 유용할 것으로 기대한다. 각 문단은 문헌과 웹에 대한 유용한 참고문헌뿐만 아니라 다른 관련된 문단에 대한 링크를 포함할 것이다.

## 통계적 검토

이 지침서는 통계학자가 연구계획서를 통계학적으로 검토하고자 할 때 고려해야 하는 체크리스트로서의 역할도 제공할 수 있을 것이다.

1. 기초적인 연구 설계: 무작위배정, 교란, 계층적인 데이터 등에 대한 정보를 포함할 필요가 있는가. 설계가 연구목적에 적합한가.
2. 데이터 유형: 대상수 산출과 제안된 통계분석 방법이 적절한지 평가할 수 있는 정보인가.
3. 대상자수: 연구 참여를 부탁할 대상자의 수와 대상수 산출로부터 모집이 예상되는 수
4. 결과변수의 총 수: 결과변수가 많아서 다중검정(multiple testing)에 대한 잠재적인 문제들을 일으킬 여지가 없는가.
5. 대상수 산출과 제안된 통계 분석이 같은 통계적 검정에 기반하고 있는지 여부: 연구자가 대상자수 산출을 보고할 때 근거가 되는 검정을 언급하고 있는가.



6. 제안된 통계 분석이 유의성 검정뿐만 아니라 신뢰구간 산출을 포함하는가.
7. 연구자가 제안된 통계 분석에 대해 필요한 통계적 전문 지식을 지니고 있는가.
8. 대상수 산출을 재현하고 체크 할 충분한 정보가 있는가.

## A. 연구 설계의 기술

### 서론

‘연구의 계획(또는 ‘대상 및 방법’)’이라고 하는 연구계획서의 부분은 연구자가 연구하려고 제안하는 바를 기술하는 곳이다. 연구의 목적과 배경은 그 연구나 개발이 합당한 이 유로 가치 있는 아이디어라고 입증하는 내용을 담을 것이다. ‘연구의 계획’에서는 연구자 가 제안하는 방법을 자세히 설명하고 실질적인 논점들의 해결방법을 제시해야 한다. 연구 계획서 심의자들은 연구가 방법론적으로 견고하고 동시에 수행가능 하다는 것과 연구자 들이 그것을 수행할 능력이 있다는 것을 확인하길 원한다(F-1). 따라서 연구수행과 관련 된 실질적인 문제들에 대해 제대로 설명하지 못한다면 검토자들은 연구의 수행가능성이 매우 낮다고 평가하게 될 것이다.

연구계획서에 구체적으로 명시되어야 할 연구목적은 연구유형에 따라 결정된다. 그러므 로 연구나 실험의 전반적인 기술에 대해 연구 설계의 표준 용어를 사용하여 초기에 연구 유형을 명확히 제시하는 것은 중요하다. 연구자들에게 익숙한 단면적 연구, 코호트 연구, 이중눈가림 무작위배정 비교임상시험 등으로 연구 설계를 구체적으로 명시하는 것이 좋 다.

### A-1 연구의 유형

#### A-1.1 연구의 유형: 관찰적 혹은 실험적 연구

연구의 유형에 있어서 가장 분명한 구분은 연구가 실험적인가(experimental) 관찰적 (observational)인가 하는 것이다. 실험적인 연구는 말 그대로 실험적인 조건으로 통제하 고 비교를 위한 대상자 군들을 구축하는 방법이다. 또한 일련의 치료법이나 개입을 수반 한다. 관찰적인 연구는 반대로 자연적인 상태에서 대상자가 관찰되는 연구이다. 비교될 대상자 군들은 육체노동자와 비육체노동자 혹은 질병이 있는 사람과 없는 사람 등으로 분류되고, 대상자에 대한 측정이나 검사는 시행될 수 있지만(예, 총콜레스테롤 측정, 질병 상태 확진) 개입이나 치료(각기 다른 운동 프로그램에 대한 환자 배정, 신약이나 위약에 대한 환자 배정 등)는 없다. 관찰적 연구에는 코호트 연구(cohort study), 환자-대조군 연 구(case-control study), 생태학적 연구(ecological study), 단면적 연구(cross-sectional study), 유병률 연구(prevalence study), 민감도와 특이도 연구(study of sensitivity and specificity) 등이 있다.

두 가지 관찰적 연구와 한 실험적 연구를 포함하는 세 가지 다른 연구에 대한 연구계획서를 가정해보자. 연구자가 퇴행성관절염에 영향을 주는 질병의 병인과 치료에 관심이 있다고 한다면 다음과 같이 연구를 계획할 수 있다:

1. 퇴행성관절염이 있는 대상자와 없는 대상자 사이에 보행 기능장애 정도를 비교한다(관찰적).
2. 퇴행성관절염이 있는 각 개인의 증상이 있는 다리와 증상이 없는 다리 간에 통증을 측정하여 비교한다(관찰적).
3. 질병이 있는 사람들을 치료군이나 비치료군으로 무작위배정하고 두 군 간에 6개월 기간에 걸쳐 통증을 측정하여 변화량을 비교한다(실험적).

연구 유형에 대한 더 자세한 내용은 역학의 원리와 응용(2005) 및 의학연구방법론(1997), 이 지침서의 B와 C에서 볼 수 있다.

#### **A-1.2 연구의 결합과 나열**

간혹 연구 설계가 각기 다른 여러 개의 연구들을 나열하여 연구계획서를 구성하는 경우가 있다. 이처럼 병합된 연구계획은 시간과 비용을 효율적으로 절감할 수 있으나 연구에 복잡성이 더해짐에 따라 연구 수행의 성공가능성은 낮아질 것이므로 연구비에 대한 부담도 늘어나게 된다. 따라서 복잡한 병합구조를 갖는 연구계획보다는 연구 설계에 대한 적절한 전문용어를 이용하여 다양한 연구들의 설계를 따로 기술하는 것이 연구 계획을 명확하게 하는 데에 도움이 되고, 연구의 성공적인 수행을 가능하게 할 것이다. 복잡한 연구를 계획하게 될 때 연구자는 복잡한 내용이 간결해 보이도록 계획서에서 기술할 수 있는지 신중히 생각해야 한다. 그렇게 하기 힘들다면, 자원 효율성은 높지만 수행가능성이 낮은 연구보다는, 단순명료한 연구가 수행가능성이 높기 때문에 연구비 수혜 목적 면에서도 더 좋은 기회를 갖게 될 것이다.

#### **A-1.3 코호트 연구**

코호트 연구(cohort study)에서 대상자는 공통적인 특성을 지니고 있는 집단(예, 같은 지리적 영역에 거주, 같은 공장에 근무, 같은 병원에 통원)으로 규정되고, 연구대상자로부터 가능한 원인 요소의 폭로에 대하여 정보가 수집된다. 인구집단은 관심 결과가 발생했는지 보기 위해 시간에 따라 추적된다. 코호트 연구에서 폭로는 질병에 대한 잠재적인 위험이고 결과는 그 질병들의 발생인 경우가 많다. 코호트 연구에 대한 더 많은 정보는 Breslow & Day(1987)에서 볼 수 있다.

#### **A-1.4 환자-대조군 연구**

환자-대조군 연구(case-control study)는 해당 인구집단(혹은 대표성 있는 표본)에서 주

어진 질병(혹은 상태)을 지닌 모든 대상자를 확인하고 질병(혹은 상태)을 가지고 있지 않은 대상자인 대조군(control group)과 비교하는 것이다. 환자군과 대조군으로부터 후향적으로 잠재적인 위험 요인에 대한 정보를 수집하여 비교한다. 환자-대조군 연구에 내재된 문제점들 중 하나는 비교할만한 대조군을 선정하는 방법이다(C-1.1). 예를 들어, 환자가 발생한 인구집단에서 질병이 없는 사람을 무작위로 추출하는 방법을 택할 수 있다. 그러나 이것은 그 인구집단에서 대상자의 목록이 존재해야 가능하다. 환자군과 대조군이 연령, 성별과 같은 변수를 비슷하게 맞추기 위해 한 명 이상의 대조군이 각각의 환자와 짝지어지는 경우도 있다. 이 경우에 짝짓기에 사용된 변수들이 질병에 영향을 미칠 수는 있지만 연구될 위험요인(risk factor)들이 질병에 영향을 미치는 경로의 일부는 아니어야 한다. 이와 같은 짝짓기의 사용은 통계분석을 복잡하게 하는 경향이 있다(C-1.2, E-5). 환자-대조군 연구에 내재된 또 하나의 문제점은 위험요인에 대한 정보가 후향적으로 수집됨으로 인해 비뚤림이 발생할 수 있다는 것이다(C-2와C-3). 이러한 문제들과 다른 문제들에 관한 정보는 Breslow & Day(1987)에서 상세히 논의된다.

### A-1.5 단면적 연구

단면적 연구(cross-sectional study)는 '2005년 인구센서스'처럼 단일 시점에서 연구 집단이나 대상자 표본을 연구하는 것이다. 즉, 표본조사(sample survey)는 단면적 연구의 한 예이다. 단면적 연구의 문제점 중 하나는 원인과 결과 중 어느 것이 먼저인지 그 순서를 명확히 설명하기가 어렵다는 것이다. 단면적 연구의 특별한 유형에는 유병률 연구(A-1.5a), 단면적 생태학적 연구(A-1.5d), 민감도와 특이도 연구(A-1.5b, A-1.5c)가 있다.

#### A-1.5a 유병률 연구

유병률 연구(prevalence study)는 대상 인구집단에서 특정한 질병/상태/특징의 유병률을 추정하는 것이다. 유병률 연구는 추정(estimation)을 일차적 목적으로 하는 표본조사이다. 이러한 유형의 연구에서 가장 중요한 사항은 대상 인구집단으로부터 대표성 있는 표본을 획득하는 것(C-4, C-5)과 연구결과에서 낮은 응답률로 인한 비뚤림이 생기지 않도록 하는 것이다(C-6).

#### A-1.5b 민감도와 특이도 추정

종종 새로운 진단법의 유용성을 평가하기 위해 민감도와 특이도가 추정된다. HIV 바이러스의 감염 여부를 판단하기 위한 새로운 ELISA 분석이 개발되었을 때 연구자들은 더 신속한 이 방법의 정확도(accuracy)를 세포배양(즉, 황금기준(gold standard))의 정확도와 비교하기를 원했다. 이를 위하여 세포배양에 따라 질병이 있는 사람과 없는 사람, 두 집단이 선정되었다. 그 다음에 ELISA 분석을 이용하여 대상자들의 양성, 음성 여부를 평가하였다. 여기서 질병을 가진 사람들 중 양성으로 판정된 사람들의 비율과 질병이 없는 사람들 중 음성으로 판정된 사람들의 비율로써 민감도(sensitivity)와 특이도(specificity)를 각각 산출하였다. 이러한 연구에서 주의해야 할 사항은 종종 연구자가 생각하는 것보다 질병을 가진 사람이 더 많이 필요할 수 있다는 것과('D. 연구대상수 산출' 참조) ELISA

검사를 시행할 때에 세포배양 결과에 대한 눈가림(blind)이 잘 이루어져야 한다는 것이다.

#### A-1.5c 민감도와 특이도의 산출 시기

특정 진단법이나 스크리닝 검사를 일차적인 스크리닝 도구로서 사용할 것을 고려하고자 한다면 황금기준에 대해 민감도와 특이도를 확인하여 결정하는 것이 필요할 것이다. 반면, 어떤 검사법이 확립된 검사를 대체할 후보라면 두 검사법 모두 황금기준과 비교해야 할 것이다. 새로운 검사법이 민감도와 특이도 측면에서 모두 확립된 검사법보다 우월하다면 새로운 검사법을 더 선호하게 될 것이다. 민감도와 특이도 중 하나만 더 우월하다면 새로운 검사법을 채택하도록 실제적인 권고를 하기 전에 위양성(false positive)과 위음성(false negative)으로 인한 비용을 고려하는 것이 필요하다.

때때로 민감도와 특이도만 산출하는 것보다는 양성예측도(positive predictive value, PPV)를 함께 산출하는 것이 더 의미 있는 경우가 있다. 예를 들어, 특수하고 다소 비용이 많이 드는 치료를 받아야 할 고위험군을 확인하는 것이 검사의 목적인 경우, 즉 제공자에게 비용이 많이 들거나 피검사자들에게 실질적으로 불리한 점이 있고, 검사상 음성인 군은 표준 치료를 계속하여 제공받는다면, 민감도와 특이도보다 양성예측도를 고려하는 것이 더 적절하다. 양성예측도는 양성으로 검사된 사람들 중 실제로 양성인 사람들의 비율이다. PPV가 낮다면 위양성이 많아지므로 위양성인 사람들의 상당수가 잘못된 진단으로 불필요하게 걱정하게 될 수도 있고, 비용이나 시간을 많이 소모하거나 불쾌감을 느끼면서 불필요한 치료를 받게 될 수도 있다. 검사대상 집단에서 유병률이 낮을 때 PPV가 낮아지는 경향이 있다.

#### A-1.5d 생태학적 연구

생태학적 연구(ecological study)는 인과성에 대한 결론을 내리려 하기보다는 개인이 아닌 집단에서 측정되는 변수들 간에 상관성을 살펴보는 것이다. 예를 들어, 같은 기간에 걸쳐 라돈 노출의 평균 수준을 추정한 여러 국가들에 대해 연간 100명당 백혈병 등록 비율의 상관관계로 백혈병과 라돈 간의 관계를 조사하는 것이다(Henshaw et al., 1990). 즉 분석의 대상은 국가이지 개인이 아니다. 이러한 유형의 연구는 특히 교란(confounding)의 영향을 받기 쉽다(A-1.6, Lilienfeld & Lilienfeld(1980) p13-15).

#### A-1.5e 측정 타당도, 신뢰도 및 일치도 연구

측정방법의 특성을 조사하는 대표적인 연구로는 타당도, 신뢰도, 일치도 연구가 있고, 이러한 연구에서 측정방법은 혈압과 같은 수치측정, 건강 상태와 같은 범주형 자료, 우울 척도처럼 설문지에 근거한 측정 등이 포함될 수 있다. 우선, 타당도(validity) 연구는 측정방법이 측정대상을 얼마나 정확하게 측정할 수 있는지를 조사하는 것이다(Bland & Altman(2002)). 타당도 연구에서 논점은 진정한 황금기준의 존재나 측정 방법에 대한 평가기준의 존재 여부, 타당도 조사 방법 등이다. 한편, 신뢰도(reliability)연구는 동일한 방법으로 같은 대상자에게 반복적으로 측정을 했을 때 유사한 결과를 나타내는 정도를 조

사하는 것이다(Bland & Altman(1996, 1996a, 1996b)). 신뢰도 연구는 동일 관찰자의 반복적인 측정 혹은 여러 관찰자들에 의한 측정과 같은 방법으로 시행될 수 있고, 시간에 따른 신뢰도나 측정 과정에서 각기 다른 부분의 측정 효과를 조사할 수도 있다. 신뢰도 연구에서 측정 대상과 관찰자의 선정은 아주 중요하다. 마지막으로, 동일한 측정대상에 대한 여러 측정방법 간의 일치도(agreement) 연구의 결과는 차후에 측정방법을 선택할 때 도움을 줄 수 있다(Bland & Altman(1986), 의학연구방법론(1997)).

#### A-1.6 교란

관찰적 연구에서 공통적으로 나타나는 문제점은 교란변수(confounding factor)의 효과로 인해 가짜 연관성이 감지되거나 진짜 연관성을 보이지 못할 수 있다는 점이다. 일반적으로 교란변수(confounder)는 연구되는 결과 변수와 인과적으로 관련성을 가지면서 설명 변수와는 비인과적 연관성을 가지는 변수이다. 따라서 질병과 잠재적 위험요인 간에 관찰되는 연관성은 실제 질병의 원인이 되는 다른 위험요인의 작용으로 나타나는 결과일 수도 있다. 이와 같은 점으로 인해 관찰적 연구로부터 인과관계에 대한 결론을 내리기가 힘들다. 관찰적 연구에서 교란이 발생하는 이유는 '비교성 있는(comparable)' 군들 간의 비교가 힘들기 때문이다. 더 많은 정보는 Breslow & Day(1980) p93-108, 역학의 원리와 응용(2005)에서 볼 수 있다.

##### A-1.6a 교란 또는 교호작용

교란(A-1.6)이라는 용어는 교호작용(interaction)과 혼동되지 않아야 한다. 교호작용은 두 변수간 연관성의 특성(크기와 방향)이 어떤 제3의 변수 값에 따라 달라질 때 발생한다. 예를 들어, 천식이 어린시절에만 여성보다 남성에서 더 흔하게 나타나는 경향이 있다면 성별과 현재 천식유무 간에 관찰되는 연관성은 연령에 따라 다르게 되며(Burr(1993)), 두 요인(현재 천식유무와 연령)간에 교호작용이 있다고 말한다. 변수들의 관계에서 이러한 교호작용을 살펴보고 기술하는 것은 중요하다. 한편, 교란변수에 대해서는 보정하거나 없애도록 설계하여, 교란변수가 관심요인과 결과변수의 연관성 감지에 영향을 주지 않도록 하여야 한다. 즉, 교란변수가 통제되었을 때 관심요인과 결과변수의 연관성의 크기와 방향이 같다는 것을 가정한다.

#### A-1.7 실험과 시험

새로운 치료법이나 개입의 효과를 평가하는 것을 목적으로 하는 실험적 연구를 시험(trial)이라 한다. 연구 대상이 동일한 의학적 조건의 사람이라면 임상시험(clinical trial)이라는 용어를 사용할 수 있다(Pocock(1983)). 연구대상이 사람, 생쥐 또는 행정적인 그룹(예, 일반의원, 진료소) 중 어느 것이든지 시험이 적용될 수 있다(A-1.8, B).

시험(예 임상시험)에서는 비교대상이 되는 그룹들이 유사한 상태가 되도록 설계할 수 있다. 비교군이 없는 시험(uncontrolled), 즉 단일 치료군만 있고 비교군으로 작용할 군이 없는 시험은 피해야 한다. 그 이유는 연구결과 개선이 생긴 경우에 새로운 치료법 때문인

지 아니면 새로운 치료법이 없이도 개선되었을 것인지 알 수 없기 때문이다. 시험에서 비교군의 필요성과 시험대상의 무작위배정 중요성에 대한 추후 논의는 A-1.8, B-3, B-5, Pocock(1983)에 있다. 교차설계시험(cross over trial)과 다른 유사한 설계에 대한 논의는 B-5.10에 제시하였다.

#### A-1.8 무작위배정 비교임상시험

무작위배정 비교임상시험(randomized controlled clinical trial)은 다른 치료법이나 개입을 비교하기 위해 설계되며, 각 군들의 구성원이 연구시작단계에서 연령과 성별과 같은 잠재적 교란변수의 분포(B-5)가 유사하게 되기 위해 연구대상자들을 무작위로 배정한다. 그렇게 해서 치료법/개입이 시행되고 결과가 추적관찰 종료시점에서 비교된다. 비교대상이 되는 군은 두 군일 수도 있고 여러 군일 수도 있다. 한 치료군과 한 대조군일 수도 있고 두 치료군일 수도 있고 두 치료군과 한 대조군일 수도 있다. 대조군은 연구대상자와 결과 평가자 모두로부터 배정상태의 눈가림을 돕기 위해 위약 처치를 받을 수도 있다. 하지만 증명된 치료법이 이미 표준적으로 사용되고 있다면 위약을 받는 대조군이나 비치료 대조군이 있는 연구는 윤리적이지 못한 것으로 간주된다(Rothman(2000) et al., F-3.3). 평가자와 연구대상자 모두 배정상태에 대해 모르게 하는 것을 이중눈가림(double-blind)이라고 한다. 단일눈가림(single-blind)은 연구대상자나 평가자 중 한쪽에 배정에 대한 정보를 주는 것을 의미한다(B-4). 앞서 기술된 시나리오 A-1.1 의 3번 항목을 살펴보면 치료를 받는 한 치료군과 치료를 받지 않는 대조군이 있는데, 시험약과 위약의 차이가 확연히 드러나는 경우에는, 연구대상자는 자신이 배정된 처치를 알게 된다. 그러나 결과를 평가하는 사람은 환자로부터 우연히 듣지 않는다면 모르는 채 유지될 수 있고, 환자에게 배정정보 누설을 금하도록 함으로써 평가자가 배정정보를 알게 될 가능성은 감소될 수 있을 것이다.

#### A-1.9 파일럿연구와 탐색적 연구

"파일럿연구(pilot study)"라는 용어를 잘못 사용하는 경우가 있다. 파일럿(pilot)은 길을 인도하는 사람이나 사물을 의미하며, 파일럿연구는 더 큰 규모의 연구를 수행하기에 앞서 소규모로 수행되는 연구를 말한다. 그러므로 파일럿연구는 그 자체만으로는 의미가 없고 반드시 규모가 더 큰 연구와 관련되어야 한다. 파일럿연구의 목적은 더 큰 연구를 쉽게 수행하도록 돕는 것이고, 데이터 수집 방법의 검증, 대상자수 산출을 위한 정보 수집 등을 위해 사용될 수 있다.

파일럿연구는 항상 그것을 활용할 본 연구가 있어야 한다. 이것은 본 연구로부터 분리한 파일럿연구 자체에 대해 계획서를 작성하지 못하거나 연구비를 신청할 수 없다는 뜻은 아니다. 그러나 대상자수나 데이터 수집 수행가능성과 관계된 파일럿연구 정보가 획득될 때까지 완전한 연구비 지원을 받지 못할 수도 있다.

파일럿연구는 질문에 대한 명확한 답을 제시할 수 없을 정도로 너무 작은 규모의 연구

를 의미하지는 않으며, 그러한 연구는 일반적으로 사회적 혜택보다는 연구자의 개인적 관심을 위해 수행되는 연구가 될 것이다.

때로는 본 연구에 대해 계획하기 이전에 예비적 단계의 아이디어가 있을 때, 소규모 연구 수행이 필요한 경우가 있다. 최종적인 연구가 많은 공동 연구자들이 참여하는 다기관 연구를 필요로 할 때, 예비적인 결과가 없이는 공동 연구자들을 모집하기가 어려울 것이다. 수행된 적이 없는 연구는 최종적인 연구를 설계할 만한 정보가 부족할지도 모른다. 그러므로 더 작은 연구에서 연구의 아이디어가 개발할만한 가치가 있는 것인지 보여줄 수 있어야 한다. 이러한 연구는 최종적인 연구가 어떻게 될지를 알지 못하는 단계이기 때문에 파일럿연구라기보다는 탐색적 연구라고 해야 한다.

## A-2 추적관찰

코호트 연구와 무작위배정 비교임상시험을 포함하는 많은 연구들은 전향적(prospective)이다. 즉 추적관찰(follow-up)기간이 있다. 그럼에도 불구하고 많은 연구자들은 추적관찰기간을 연구계획서에 명시하지 않은 것을 보게 된다. 전체 추적관찰기간에 대한 정보 없이 측정이 3개월마다 반복된다고 언급하면 환자 당 측정되는 횟수를 알 수 없게 된다. 추적관찰기간에 대한 정보는 프로젝트의 수행능력을 평가할 때 결정적일 수 있다. 예를 들어, 어떤 특정 암에 대한 치료법의 무작위배정 비교임상시험에서 치료하지 않은 군의 재발률이 80%라고 가정하여 대상자수를 산출하였다고 하자(D-8.2). 여기서 연구기간이 5년인 경우에 대하여 연구대상수를 산출하였다면, 80%라는 수치는 5년에 걸친 재발과 관련이 있다. 그러나 제안된 추적관찰 기간이 단지 2년뿐이라면 연구 결과는 절대적으로 검정력이 부족하게 될 것이다. 추적관찰 기간은 시간에 따라 소멸하기 쉬운 환자교육프로그램과 같은 개입의 영향을 평가하는 시험에서도 중요하다. 이런 상황에서 개입기간 직후에만 결과를 평가하는 것은 그리 유용하지 않으므로 추적관찰을 통한 평가가 이루어져야 한다.

## A-3 연구대상

연구 질문에 답하기 위해서는 연구 대상이 어느 모집단을 대표하며 연구가설 검정에 적합한 대상인지 파악하는 것이 중요하다. 예를 들어, 젊은 사람에게 많이 발생하는 특정 질병에 대하여 중년을 대상으로 연구하는 것은 적절하지 않다. 연구 대상이 어떻게 선정될 것인지 파악하는 것 또한 중요하다. 예를 들어, 모집단에서 무작위 추출된 사람들을 연구대상으로 할 수도 있고 또는 특정 기간에 진료소를 방문한 모든 환자들을 대상으로 할 수도 있다. 그리고 선정/제외 기준을 구체화하는 것도 중요하다. 연구자는 얼마나 많은 대상자가 연구에 필요하고 연구에 동의할 것으로 예상되는지 언급해야 한다. 대상자수는 연구자가 접근할 수 있는 수가 아니라 연구 참여에 동의한 환자의 수라는 것을 명심해야 한다.

## A-4 변수 유형

제안된 연구에서 조사될 결과변수와 설명변수의 유형과 측정 척도를 구체화하여 기술하는 것이 중요하다(A-4.1, A-4.2). 다음 정보는 적절한 대상자 수 산출방법뿐만 아니라 통계적 분석의 특성을 결정하는 데에 도움을 줄 것이다.

### A-4.1 측정의 척도

1. 구간척도(interval scale): 데이터가 자연적인 순서를 가지고 있고 값들 사이의 간격이 의미가 있다. 예: 체중, 신장, 출산수.
2. 순위척도(ordinal scale): 데이터가 자연적인 순서를 가지지만 값들 사이의 간격은 의미를 가질 필요가 없다. 예: 정신과적인 척도.
3. 명목척도(nominal scale): 범주가 자연적인 순서를 가지지 않는 범주형 데이터. 예: 성별(남/여).

### A-4.2 자료의 유형

정량적 자료(quantitative data): 간격척도로 측정되는 자료

1. 연속형 자료(continuous data): 주어진 범위에서 모든 가능한 값을 취할 수 있는 연속적인 값을 가진 변수. 예: 체중, 신장.
2. 이산형 자료(discrete data): 주어진 범위에서 단지 한정된 숫자 값만을 취할 수 있고, 보통 자연수나 정수로 표현되는 변수. 예: 출산수.

정성적 자료(qualitative data): 자연적인 순서가 있을 수도 있고 없을 수도 있는 범주로 표현되는 자료. 예: 명목척도, 순위척도.

### A-4.3 자료 수집의 방법

연구의 질은 데이터의 질에 따라 크게 좌우된다. 따라서 연구계획서 심사자는 연구자가 정보 수집을 어떻게 계획하였는지 확인해야 한다. 설문지를 사용할 계획이라면, 우편발송이나 직접면접 등 어떤 방법으로 설문지를 피조사자에게 전달할 것인지 명확히 제시하여야 한다. 면접을 직접 수행하는 경우는 데이터 수집의 완결성에 도움이 될 수 있지만, 면접원이 적절하게 눈가림이 되지 않는다면(예를 들어, 환자-대조군 연구에서 환자/대조군 상태, 임상시험에서 치료군) 비뚤림(bias)이 생길 수 있다. 연구자가 주치의 노트나 병원 기록지와 같은 기록으로부터 정보를 추출하는 것을 제안할 때는 수동으로 혹은 데이터베이스 검색으로 혹은 두 가지 방법을 모두 사용하여 수행할 것인지 등이 명시되어야 한다. 이 경우에도 추출하는 사람이 적절하게 눈가림이 되지 않는다면 비뚤림이 발생할 수 있다(C-2). 연구자는 선택된 수집 방식이 완전하고 정확한 정보를 얻기에 충분한지 또한 스스로에게 자문해 보아야 한다. 예를 들어, 병원전산자료 등에서 검색하는 것만으로는



인과성에 대한 완전한 정보를 얻을 수 없을 것이다. 폐기능 평가 자료를 산출하고자 할 때, 폐활량계(spirometer)로 한 번 측정하는 것만으로 신뢰성 있는(A-4.4b) 폐기능 평가 자료를 얻지는 못하므로 연속적으로 세 번 측정한 것 중에서 최대치를 이용하는 경우가 많다.

#### A-4.4 타당도와 신뢰도

연구자는 제안된 측정 방법의 타당도와 신뢰도에 대해 제시해야 한다. 이것은 제안된 방법이 비교적 새로운 방법이거나 특정 분야 외에는 흔히 사용되지 않는 것이라면 더욱 중요하다. 더 많은 정보는 A-4.4a, A-4.4b, A-4.4c, Altman(1991)과 Bland(2000), 의학연구방법론(2005)에서 볼 수 있다.

##### A-4.4a 타당도

타당도란 측정하고자 하는 방법이 연구자가 가정하는 것을 정확하게 측정하는지 보는 것이다. 예를 들어, 어떤 순위척도에 대한 자기-효용성(self-efficacy)을 측정하기 위해 설계된 설문지가 있다고 가정했을 때 이 설문지로 자기-효용성을 정확하게 측정할 수 있는가? 심지어 타당성이 이미 입증된 경우에도 이 적절성을 증명할 수 있는 적절한 환경에서 측정되었는지 확인해야 한다. 예를 들어, 성인을 대상으로 사용하는 것이 타당하다고 입증된 방법이 어린이에게 사용될 경우에는 타당하지 않을 수 있고, 특정 국가에서 평가된 타당도를 다른 나라에 적용하면 안되는 경우도 있다. 간혹 연구자들은 이전에 타당성이 입증된 설문지 점수를 참고하여 개정된 버전을 사용하고자 하기도 한다. 이 때 이들 개정판이 타당도에 어떤 영향을 주는지 검토자는 확인해보아야 한다. 또한, 자료처리를 쉽게 하기 위해 점수를 산출하는 질문 수를 줄이는 경우에도 마찬가지로 타당도에 미칠 영향을 평가해야 한다. 더 많은 정보는 Bland & Altman(2002)에서 볼 수 있다.

##### A-4.4b 반복성(검사-재검사 신뢰도)

반복성(repeatability)은 한 대상에 대한 단일 측정값이 그 대상에 대한 평균값을 얼마나 정확하게 추정하는가를 의미한다. 따라서 측정의 반복성은, 같은 대상자에 대해 반복된 측정의 표본을 이용하여 산출할 수 있는 대상 내 표준편차(within-subject standard deviation)에 따라 다르다. 반복성 계수는 단순히 대상 내 표준편차에 2.83을 곱한 것이고, 같은 대상에서 두 번 연속측정 간에 발생할 수 있는 최대 차이의 추정치이다(Bland & Altman(1986, 1996, 1996a and 1996b)).

##### A-4.4c 평가자간 신뢰도(평가자간 일치도)

평가자의 역할이 중요한 측정 방법에 대해서는 평가자간 신뢰도 또한 고려되어야 한다(Altman(1991) p403-409). 즉, 같은 대상을 다른 관찰자/평가자가 평가했을 때 어느 정도의 차이가 측정될 수 있을 것인가? 어느 정도 일치할 것인가? 동일한 관찰자가 계속 평가하지 않고 평가대상에 따라 관찰자가 다르다면, 연구결과에 심각한 비뚤림이 발생할 것이다. 그러나 간혹 다기관 임상시험처럼 여러 명의 관찰자가 평가해야 하는 경우가 있다.

이러한 경우에 1) 사전 훈련으로 측정자간 일치도를 향상시키고, 2) 같은 대상자에 대한 전/후 측정을 할 때 동일한 평가자가 평가하도록 하고, 3) 임상시험에서 평가자에 대해 군들의 균형을 맞추고 모든 측정에서 눈가림이 되도록 하고(Pocock(1983) p45-48), 4) 관찰적 연구에서 관찰자가 분석에서 잠재적 교란변수로서 보정될 수 있도록 각 대상을 평가한 관찰자를 기록해두는 것이 중요하다.

## B. 임상시험

### B-1 연구대상과 선정제외기준 기술

연구계획서는 어떤 환자군을 대상으로 연구를 하는지 명확히 제시하여야 한다. 예로, 고혈압연구에서 어떤 범위의 혈압을 가지는 환자들을 대상으로 하는지 명시하는 것처럼, 연구에 등록될 대상의 기준을 구체적으로 기술하여야 하고, 이것을 선정제외기준이라고 한다. 경/중증의 고혈압을 가진 노인을 대상으로 실시된 한 고혈압 관련연구는 네덜란드 로테르담 근교에 거주하는 노인 외래환자로 구성된 코호트(이하 로테르담 코호트)에서 연구대상자들을 등록하였다(Geleijnse et al.(1994)).

로테르담 코호트의 모든 구성원들은 1990년~1992년에 혈압을 측정하였고 측정당시의 혈압 이 140/85 mmHg 이상이고 항고혈압제를 투여하지 않은 55-75세의 남녀에게 편지나 전화로 재측정을 요청하였다. 연구에 등록되는 대상자로 적절하기 위해서는, 일주일간의 간격을 두고 2번 측정된 혈압이 SBP 130 mmHg 이상, DBP 70 mmHg 이상이면서, SBP 140~220 mmHg 혹은 DBP 85~110 mmHg이어야 했다. 이를 요약하면, 선정기준(inclusion criteria)과 제외기준(exclusion criteria)은 다음과 같았다.

#### 선정기준

로테르담 코호트의 구성원

등록당시 연령이 55-75세

1990년~1992년 혈압측정 시에 치료되지 않은 SBP 140mmHg 이상 혹은 DBP 85mmHg 이상

근래 일주일간의 간격을 두고 2번 측정된 혈압이 SBP 140~220 mmHg 혹은 DBP 85~110 mmHg 이면서 SBP 130 mmHg, DBP 70 mmHg 이상

#### 제외기준

심근경색증의 과거병력

협심증의 과거병력

당뇨병

신장기능 손상 (serum creatinine concentration > 200 mmol/L))

처방에 따른 제한된 염분 섭취

한편, 임상시험심사위원회(Institutional Review Board)에서는 연구자가 취약한 환경에 있는 이들을 보호하기 위하여 어떠한 규정을 마련하였는지를 확인한다. 연구자들은 이러한 사람들을 연구대상으로부터 제외함으로써 이들을 보호하기 위한 규정을 만들지 않아도 되겠지만, 이들의 치료가 가능하다면 등록 또한 가능해야하므로, 임의로 연구등록을 막는 것은 적절하지 못하다.

## B-2 치료법 기술

연구계획서에 치료법을 명확히 기술하는 것은 중요하다. 통계학자를 포함하여 연구계획서를 읽는 모든 사람들이 임상적으로 자질을 갖춘 것은 아니므로 기술된 치료법은 이러한 사람들도 읽고 이해할 수 있도록 명확하고 평이한 서술로 쓰여야 한다.

일반적으로 의학연구는 타 영역의 사람들에게는 익숙하지 않은 시험설계를 포함할 수 있다. 예로, 연구자 혹은 시험 참가자들이 피험자들의 배정된 치료법(treatment)을 아는 것이 연구결과에 영향을 미칠 수 있는 경우 이를 방지하기 위하여 눈가림법(blinding)을 사용하게 된다. 하지만, 눈가림과 같은 시험설계를 잘 알지 못하는 사람들이 이를 이해하기는 힘들기 때문에 연구계획서 초안이 작성되면 타 분야의 사람에게 검토를 부탁하여 치료법이 쉽게 이해가 되는지를 확인하여야 한다.

## B-3 대조군의 필요성과 선택방법

시험(trial)이나 실험(experiment)의 핵심은 평가하려는 치료법을 적용하고 그 결과를 관찰하는 것이다. 어떤 형태의 개입(intervention)이 가해지지 않는다면, 이는 시험이 아닌 관찰적 연구(observational study)가 된다. 임상시험(clinical trial)에서는 개입이 어떠한 방식으로 질병의 진행을 개선시키는데 관심이 있다. 이때 개입은 내과적, 외과적 치료와 같이 직접적인 방법일 수도 있고 혹은 일차 진료기관에 치료 지침을 제공하거나 건강교육프로그램을 시행하는 것처럼 직접적이지 않은 방법일 수도 있다.

대부분의 연구에서 치료의 효과여부를 살펴보기 위하여 대조군(control group)을 필요로 한다. 페니실린의 경우, 처음 발견되었을 때 페니실린 투여에 의한 효과는 과거에 보였던 그 어떤 것보다도 확연히 달랐기 때문에 대조군 없이 과거 사례의 일반적인 경험과 비교하여 연구가 시행되었으나, 후에 보다 경미한 감염에 적용되기 시작하였을 때에는 비교시험(controlled trial)이 수행되었다(Pocock(1983)). 그러나 대부분의 경우 평가할 치료의 효과가 초창기 페니실린의 경우만큼 크지는 않으므로 대조군 없이는 치료의 효과를 제대로 평가하기는 힘들다. 따라서 새로운 치료법을 받는 피험자군과 새로운 치료법을 받지 않는 피험자군의 비교가 필요하고 전자를 시험군(intervention group), 후자를 대조군(control group)이라 한다.

대조군은 질병의 중증도와 예후, 장소, 시기 면에서 새로운 치료를 받는 군과 유사한 분포를 가져야 하고, 관심이 되는 치료를 제외한 나머지는 모든 면에서 시험군과 비슷해야 한다. 이를 실현할 수 있는 가장 신뢰할 만한 방법은 무작위 할당(random allocation)이나 최소화법(minimization)이다. 교대로 치료법을 할당하는 것과 같은 방법과 같이 엄밀하게는 무작위적인 배정방법이 아닌 설계는 피해야 하고, 다른 시기 혹은 다른 장소의 피험자 등록과 같이 두 군의 비교성을 보장할 수 없는 방법이 제안서에 있을 경우, 그 방법이 타당한 이유를 정확하고 명확하게 기술하는 것이 필요하다.

시험을 수행할 때 연구자는 어떤 치료법의 무엇을 검정하려고 하는지 분명한 아이디어를 가져야 한다. 제안서에는 시험의 목적과 시험의 필요성이 기술되어야 한다. 평가할 치료법뿐만 아니라 대조군이 어떤 치료를 받을지는 논의거리가 된다. 대조군은 어떤 식으로라도 치료를 받아야 하는가? 만약 현재 가능한 치료가 없다면, 대조군은 실제 치료를 받지 않을 것이고 눈가림(blindness)을 유지하기 위해 가짜치료(dummy treatment)를 사용할 수도 있다. 만약 기존의 치료법이 있는 경우라면, 시험군에는 기존 치료에 새로운 치료가 가해지고, 대조군에는 기존 치료 혹은 기존치료에 새 치료의 적절한 가짜치료가 가해질 수 있을 것이다. 현재 헬싱키 선언(Declaration of Helsinki)에서 다음과 같이 제시하고 있다.

“새로운 방법의 이익, 위험, 부담과 유효성(efficacy)은 현존하는 최상의 예방, 진단, 치료방법과 비교하여 시험되어야 한다. 증명된 예방, 진단 혹은 치료방법이 없는 경우에도 대조군에게 위약(placebo) 혹은 무치료(no treatment)를 설정하지 않아야 한다는 것은 아니다.” (29 절)

이후 아프리카의 HIV 연구처럼 현존하는 최상의 치료가 너무 비싼 경우에 관련된 임상시험을 수행하고자 했던 연구자(Ferriman 2001)들과 제약관련 연구자, 약물관련 규제자들 (Tollman et al.(2001))은 항의를 하기도 하였다. 세계의학협회(World Medical Association, WMA)는 다음과 같은 성명을 발표했다.

“세계의학협회는 개정된 헬싱키선언의 29절이 해석의 혼란을 야기할 수 있음을 우려하는 바이다. 따라서 위약비교연구(placebo-controlled trial)에는 주의를 기울여야 하고, 일반적으로는 현존하는 증명된 치료가 없는 경우에 적용되어야 한다는 입장을 재확인한다.” 그러나, 다음과 같은 상황에서는 입증된 치료가 있더라도 위약비교연구를 윤리적으로 허용할 수 있다.

1. 필연적이며 과학적으로 옳은 방법론적 이유로, 예방, 진단, 치료방법의 유효성과 안전성을 결정하는 데 필요한 경우
2. 예방, 진단 혹은 치료방법이 경미한 상태를 대상으로 하고, 위약(placebo)을 투여 받

는 피험자는 중대하거나 비가역적인 손상 등의 어떠한 추가적인 위험에 노출되지 않을 경우

‘적절한 윤리적, 과학적 검토(ethical and scientific review)의 필요성’에 대한 조항을 포함하여 헬싱키 선언의 모든 조항들은 반드시 지켜져야 한다.”

이 성명은 일부 비평가(Bland(2002a, 2002b))에게는 만족할 만하지 못했고, 여전히 논쟁거리이지만, 위약비교 형태의 모든 연구는 어떠한 이유로 대조군의 위약치료 선정이 정당한지 계획서에 명시해야 할 것이다.

때로는, 둘 이상의 대조 치료법(control treatment)의 선정이 가능하다. 예로, 새로운 고혈압 치료제를 평가하고자 하는 경우, ACE-inhibitor, beta-blocker, 이뇨제, 염분제한 식이, 운동 중 하나 혹은 이 중 둘 이상의 조합을 대조 치료법으로 하여 비교할 수 있다. 부적절한 대조 치료법을 계획하거나 최상의 대조 치료법이 아닌 방법을 계획하는 임상시험은 비판을 당할 수 있다. 대조 치료법을 선택하는 임상시험에서는 연구계획서에 선택된 대조 치료법에 대한 정당성을 명시해야 한다.

## B-4 눈가림

### B-4.1 이중눈가림과 단일눈가림 설계

응답에서의 비뚤림(bias)은 연구 참여자의 잠재의식에 따라 일어날 수 있다. 환자의 반응은 그가 어떤 치료를 받고 있는지에 대한 인지여부, 특정 치료가 유익한지 해로운지에 대한 믿음, 임상의의 마음에 들고 싶은 욕구 등에 따라 영향을 받게 된다. 특히 활성약(active drug)과 비활성약(inert drug)을 비교하는 연구에서 피험자가 어떠한 치료를 받는지에 대해 관찰자가 알게 된다면 관찰 결과에 영향을 미칠 것이다. 이러한 이유로, 피험자가 어떤 치료를 받고 있는지 피험자 본인과 평가자 모두 모르게 하는 이중눈가림법(double blind)이 선호된다. 때로는 연구 참여자 중 한쪽만 치료에 대해 눈가림이 되는 단일눈가림(single blind)만이 가능한 경우가 있는데, 대부분이 피험자가 받는 치료에 대한 임상의의 눈가림이 불가능하여 환자에게만 눈가림이 가능한 경우가 많다. 한 예로, 치료가 수술의 형태인 경우가 될 수 있다. 또한, 환자와 임상의 모두에게는 눈가림이 불가능하지만, 피험자가 어떤 치료를 받았는지 모르는 제 3의 관찰자가 X-ray 결과를 평가하는 경우와 같이 평가의 눈가림이 가능한 경우가 있다. 일반적으로 눈가림법을 적용하는 연구에서는 연구 참여자에게 어떻게 눈가림이 이루어지고 유지될 것인지 자세히 기술하는 것이 중요하다.

연구에 피험자를 등록시키는 사람이 다음에 등록되는 환자가 어떤 치료를 받게 되는지 모르게 하는 것 또한 바람직하다. 이에 대한 구체적인 방법은 B-5 장에 제시되어 있다. 요약하면, 이중눈가림연구는 피험자와 평가자 모두, 단일눈가림연구는 둘 중 한 쪽에 한

해서 치료에 대한 식별이 불가능하도록 유지하는 것이다.

#### B-4.2 위약

새로운 치료법을 치료가 없는 경우와 비교하기 위한 이중눈가림연구를 시행하는 경우, 가짜 약(dummy pill), 즉, 위약(placebo)을 주어야 한다. 위약의 처치는 피험자가 치료에 대해 식별 불가능하도록 하고, 특정치료를 받는 것을 앎으로써 피험자의 상태가 호전되었다고 느끼는 심리적인 효과를 방지하게 한다. 이러한 치료에 대한 심리적인 반응을 위약 효과(placebo effect)라 한다. 경우에 따라서는 위약효과가 매우 큰 경우도 있다. 예로, 3가지 활성약이 위약과 비교되는 진통제 시험에서 각 약이 4가지 색으로 제조되었을 때 전반적으로는 활성약이 위약보다 효과가 있었으나 이상하게도 빨간색 위약은 활성약만큼 유효하게 관찰되었다(Huskisson(1974)).

위약은 외양과 맛이 활성약과 동일해야하고, 약물학적으로 비활성이어야 한다. 위약의 사용은 새로운 치료의 효용성과 부작용(side effect)을 평가 가능하게 한다. 위약은 비약물 시험에서 사용될 수 있으나, 비윤리적일 수 있다. 예로, 백신연구에서 식염주사가 사용될 수 있으나 윤리적으로 허용되지 않을 수 있다.

때로는 같은 외양을 가지기 힘든 두 약을 비교하는 경우 피험자에게 배정된 활성약과 함께 다른 군의 활성약과 같은 외양을 가지는 위약을 주는 이중위약(double dummy)을 이용하여 눈가림을 유지할 수 있다. 예를 들어, 정제와 외용제를 비교하는 경우 각각의 피험자는 활성화 정제와 위약 외용제 혹은 위약 정제와 활성화 외용제 중 하나를 받게 된다.

#### B-5 무작위배정

##### B-5.1 무작위배정

무작위배정(randomization) 혹은 무작위적인 할당(random allocation)은 피험자의 특성이 할당될 그룹에 영향을 미치지 않게 각각의 그룹으로 피험자들을 나누는 기법이다. 이는 각 피험자가 할당될 그룹이 우연에 의해 결정되고 따라서 그룹 간에 생기는 차이는 모두 우연에 의한 것으로 간주할 수 있게 된다는 것이다. 임상시험에서 무작위배정은 피험자가 어떤 치료를 받게 될지를 결정하는데 사용될 수 있다. 이를테면, 새로운 치료와 기존치료를 비교하는 시험에서 1:1 무작위배정을 하는 경우 각각의 피험자가 새로운 치료 혹은 기존치료를 받을 확률은 같다.

##### B-5.2 무작위배정이 적용되는 시험

다른 처치를 받는 그룹을 비교하고자 하는 실험적 연구(experimental study)에서 무작위배정을 사용할 수 있다. 이때 연구대상은 인간에 국한되는 것이 아니라 동물, 다른 생물학적 혹은 유기체 단위도 될 수 있다. 무작위배정이 적용되는 전형적인 예로는 피험자

가 두 가지 치료 중 하나를 받게 되는 임상시험이 있다. 또한 개개인을 둘 이상의 그룹에 할당하거나 피험자들의 그룹을 서로 다른 치료군으로 할당하고자 할 때 사용될 수 있다. 예를 들어, 진료기관 전체를 2개의 치료 중 하나에 할당하는 집단 무작위배정(cluster randomization)이 있다.

### B-5.3 무작위배정의 필요성

임상시험에서 무작위배정이 선호되는 이유는 3가지로 나누어 볼 수 있다.

첫째, 비뚤림(bias) 없이 치료군 간의 차이를 평가하기 위해서 이다. 연구자는 치료군 간에 관찰되는 모든 차이는 치료군의 효과차이로 인한 것이라고 결론짓기를 원한다. 즉, 피험자 특성의 차이가 치료차이와 교란되기를 원하지는 않는다. 무작위배정이 없는 치료의 비교는 고의적이던 그렇지 않던 간에 특정 피험자들에게 특정 치료를 받게 함으로써 비뚤린 결과를 초래할 수 있다. 무작위할당이 주어지는 치료 이외에 그룹들을 완전히 동일하게 할 수는 없지만, 그룹간의 차이가 있다면 그러한 차이는 우연에 의해서 발생한 것이라 할 수 있다. 둘째, 무작위배정은 치료의 비교에서 발생하는 비뚤림(bias)을 줄이기 위해, 연구자와 피험자가 치료형태를 알 수 없게 한다. 셋째, 무작위배정은 모집단의 임의 표본(random sample)인 치료그룹이 추출되도록 하여, 확률론에 근거한 표준적인 통계적 검정을 타당하게 한다.

### B-5.4 무작위가 아닌 배정방법

어떤 연구는 새로운 치료를 받는 현재 환자와 기존치료를 받은 과거 환자를 비교한다. 이러한 경우, 환자들을 무작위배정하는 것은 불가능하다. 과거 대조군(historical control)은 현재 환자와 많은 면에서 다를 수 있고, 새로운 치료를 받는 현재의 환자와의 비교에서 비뚤림이 생길 수 있다(B-3 참조).

다른 일반적인 접근방법으로 계통적 할당(systematic allocation)이 있다. 교대로 이루어지는 할당(예, A B A B...)이나 생일, 연구에 참여하는 날(짝수 날이면 A, 홀수 날이면 B)을 이용하는 것이 예가 될 수 있다. 연구비 신청서나 저널의 논문에서는 배정이 무작위로 이루어졌다고 하였으나 후에 피험자가 치료군에 교대로 배정된 경우를 볼 수 있다. 이러한 설계는 원칙적으로는 비뚤리지 않았으나, 무작위배정절차에 접근권한이 있는 사람이 할당을 변화시킬 수 있는 여지가 있기 때문에 문제가 발생할 수 있다. 이러한 이유로 다른 방법이 있는 한, 계통적인 할당은 권장되지 않는다.

### B-5.5 무작위배정 방법

가장 뚜렷하고 단순한 무작위방법은 동전을 던지는 것과 같은 물리적인 방법을 사용하는 것이다. 즉, 피험자가 시험에 등록될 때마다 동전을 던져 무작위배정이 이루어지고 이 결과를 피험자 정보지(patient information sheets)에 기술한다.

그러나 일반적으로 피험자가 등록될 때마다 동전을 던지는 방식과 같은 무작위배정을 피해야 하는데, 가장 큰 이유는 점검기록(audit trail)을 남기지 못하기 때문이다. 점검기록이 존재하지 않으면 무작위 할당이 제대로 이루어졌는지를 확인할 길이 없다. 예를 들어, 연구자들이 할당된 배정에 불만족하였더라도 동전을 다시 던지지 않았음을 보장할 수 있는 방법이 없다. 따라서 피험자가 시험에 등록되기 전에 미리 작성한 배정표를 토대로 무작위 할당이 결정되어야 한다. 무작위배정표는 연구에 참여하는 피험자를 볼 수 없는 제 3자에 의해 만들어져야 하고, 환자를 등록하는 사람들은 배정에 대해 눈가림이 되어야 한다.

대규모 연구에서의 무작위배정의 경우, 동전을 던지는 대신 수리적인 절차에 의해 만들어진 난수를 사용할 수 있고, 일련의 난수를 만들기 위해 컴퓨터 프로그램에 의해 만들어진 난수표(tables of random numbers)를 사용할 수 있다. 예로, 주어진 난수표에서 주사위를 던지거나 다른 유사한 방법으로 임의의 시작점을 선택한 후, 다음 난수가 홀수이면 새로운 치료 짝수이면 기존치료와 같이 배정표를 만들 수 있다. 하지만, 컴퓨터의 접근성이 좋아짐에 따라 이러한 절차 대신 컴퓨터만을 이용하여 무작위배정 수행을 위한 프로그램을 작성할 수 있으며, 교재나 웹을 통해서 무작위배정을 위해 작성된 통계패키지의 프로그램 혹은 무작위배정 사이트를 이용할 수도 있다. 의학연구협력센터 홈페이지(<http://mrcc.snu.ac.kr>, <http://mrcc.snuh.org>)에서도 실시간으로 운용할 수 있는 무작위배정 구현을 지원하고 있다. 이러한 프로그램들에서는 첫 피험자 등록 전에 무작위배정표를 작성하여 출력하거나 피험자가 한명씩 등록되면서 무작위배정표가 작성되며, 무작위배정 방법을 연구계획서에 자세히 기술하는 것은 매우 중요하다.

무작위배정표(randomization list)는 연구의 등록에 관여하지 않은 누군가에 의해 준비된 후, 연구자들이 이용할 수 있도록 제공되어야 한다. 무작위배정은 물리적으로 가까운 지점뿐만 아니라 장거리의 경우에도 피험자 등록시점에 전화를 통한 방법 등을 통해 시행될 수 있다. 병원현장에서 무작위배정을 하는 한 방법으로 무작위배정순서를 봉투에 넣는 방법이 있을 수 있다. 이 경우, 연구자가 봉투 안의 내용물을 읽기 위해 램프근처에 봉투를 비쳐본 사례 등이 있으므로 봉투는 내용이 보이지 않도록 불투명하게 하는 것이 중요하다. 같은 이유로 피험자를 등록하는 사람이 다음 봉투를 올바르게 취하기 위해 봉투에 일련번호가 매겨져야 한다.

연구자들이 다음에 등록될 환자가 어떠한 치료를 받을지 알게 해서는 안되므로, 무작위배정표는 연구자에게 공개해서는 안된다. 무작위배정이 연구자에게 공개된 경우 명백한 비뚤림을 야기하고 치료군 간의 결과는 과장되게 큰 차이를 보인 경우가 빈번히 발생하여 왔다.

전화에 의해 이루어지는 장거리 배정(long range allocation)은 큰 규모의 연구와 다기관 연구(multi-center trial)의 경우에 적절하고 사무실에 전화를 받을 누군가가 있어야 한



다. 이는 연구되는 질환에 따라 정규 근무시간이나 24시간이 될 수 있다. 대개 연구자들은 이러한 무작위배정 운용을 하기 어려우므로, 24시간 전화 서비스(이 서비스는 때때로 컴퓨터에 의해 작동되기도 한다.)를 제공하는 상업적인 시험 사무소를 이용하는 경우가 많다.

무작위배정 기록을 유지하는 것은 좋은 생각이다. 각 그룹에서 나이, 성별, 중요한 예후 인자와 같은 변수의 분포가 유사한지 확인하는 것은 중요하며, 특히 제 3의 집단에서 전화로 무작위배정을 제공할 때 중요하다.

### B-5.6 블록 무작위배정

항상 그룹 간 피험자 수를 유사하게 유지하고자 하는 경우에는 블록 무작위배정(block randomization)을 사용할 수 있다. 한 예로, A와 B, 2가지 치료가 있고, 피험자를 블록크기 4로 배정하고자 한다고 가정하면 치료 A에 2명, 치료 B에 2명씩을 할당하는 방법은 총 6가지가 있다.

1. AABB    2. BBAA    3. ABAB    4. BABA    5. ABBA    6. BAAB

만약 6가지 방법의 조합을 사용하면, 그룹간의 피험자의 수는 어떤 등록시점에서나 2이상의 차이는 나지 않는다. 배정순서는 난수표를 이용하여 무작위로 블록을 선택함으로써 만들어진(예를 들어, 1은 AABB, 2는 BBAA와 같이 블록의 번호를 매기고, 1-6 이외의 난수는 무시한다). 블록 배정 또한 컴퓨터를 이용하여 이루어질 수 있다.

임상시험에서 연구자들이 피험자가 어느 치료에 배정될지 미리 추론하는 것을 피하기 위한 최선의 방법은 연구자들이 무작위순서가 어떻게 만들어졌는지 모르게 하는 것이다. 이러한 이유로 큰 규모의 시험에서는 큰 블록크기를 사용하여 순열을 추측하기가 힘들게 할 수 있다. 이러한 순열을 만들기 위해서는 컴퓨터를 이용하는 것이 필요하다.

### B-5.7 층화 무작위배정

무작위배정의 목표는 서로 다른 치료를 받는 피험자군 사이에 예후에 영향을 미칠 수 있는 특성에 관한 분포를 가능한 유사하게 하는 것이다. 이를테면, 예후는 종종 나이와 관련되어 있으므로, 연구자는 그룹 간 유사한 연령분포를 가지기를 원한다. 그러나 작은 규모의 연구의 경우 무작위배정이 그룹 간 균형을 보장하지는 않는다. 우연에 의해 차이가 발생할 수 있고, 이는 결과의 신뢰도와 관련하여 의구심을 야기할 수 있다. 이에 대한 해결책의 하나는 시작시점부터 예후에 영향을 많이 미치는 변수에 대해 층화 무작위배정(stratified randomization)을 하는 것이며 또 다른 방법은 최소화법(minimization)을 사용하는 것이다.

층화 무작위배정은 각각의 하부그룹(층)에 대해 무작위배정표를 작성하여 각 층 내에서

치료군의 피험자 수를 유사하게 하는 것이다. 예로 유방암에 대한 2가지 대체치료에 관한 연구에서 폐경상태를 고려하면, 폐경기 전의 여성군과 후의 여성군에 대해서 독립적인 2개의 무작위배정표를 작성하고 봉투에 넣어 봉합한 상태로 보관한다. 이때 각 층 안에서 치료간의 균형을 유지하기 위해 부가적으로 블록을 사용할 수 있다(B-5.6 참조). 층화 무작위배정은 2개 혹은 그 이상의 층화변수로 확장될 수 있다. 하지만 층의 수가 많아지면 하부그룹의 규모가 너무 작아지므로 적은 수의 층을 이용하는 것이 좋다.

## B-5.8 최소화법

중요한 예후 인자가 많은 작은 규모의 연구에서는 무작위배정으로는 그룹간 적절한 균형을 유지할 수 없다. 또한 층화 무작위배정(B-5.7참조)은 적은 피험자를 중요한 변수 모두에 의해 층화시킬 수 없으므로 실현 불가능할 수 있다. 이러한 경우, 최소화법(minimization)이라는 기법을 이용하여 균형을 유지할 수 있다. 이는 다음 환자가 연구에 등록될 무렵 그룹 간 전체적인 불균형을 최소화하는 치료를 배정한다는 아이디어를 기반으로 한다. 연구계획서에서는 어떠한 예후인자를 사용할 것인지 어떻게 그룹화 될 것인지 명시하는 것이 중요하다. 예를 들어, '연령을 예후인자로 사용하겠다.'고 하는 것만으로는 충분하지 않다. '50세 미만과 50세 이상으로'와 같이 실제 연령군이 제시되어야 한다.

간략히 말하면 최소화법은 다음과 같이 수행된다. 처음 피험자가 A 혹은 B에 무작위배정된다. 다음 피험자가 등록되고 그들의 예후인자가 기재되면, 그 시점에서 전체적인 균형이 최적화되게 하는 군으로 배정된다. 예로, 한 연구에서 16개의 의원이 시험군과 대조군으로 배정되고, 그룹 간 균형을 이루어야 하는 3개의 변수는 다음과 같았다.

의원에서의 의사의 수  
 의원에서의 환자의 수  
 장기간 정신병 환자의 수

이를 다음과 같이 그룹화 한다.

의원에서의 의사의 수 : 3, 4, 5, 6명  
 의원에서의 환자의 수 : 8,600명 미만/이상  
 장기간 정신병 환자의 수 : 25명 미만/이상

처음 의원은 다음과 같은 특성을 보였다.

의원에서의 의사의 수 : 4명  
 의원에서의 환자의 수 : 8,500명  
 장기간 정신병 환자의 수 : 23명

처음에는 두 치료군 모두 등록된 피험자가 없으므로, 첫 의원의 경우 두 그룹 중 무작위적으로 배정되어 시험군에 배정되었고, 다음과 같은 최소화 변수에 대한 표가 만들어졌다.

	시험군(Intervention)	대조군(Control)
3, 4 명의 의사	1	0
5, 6 명의 의사	0	0
8,600 명 미만의 환자	1	0
8,600 명 이상의 환자	0	0
25명 미만의 정신병 환자	1	0
25명 이상의 정신병 환자	0	0

두 번째 의원은 다음과 같은 특성을 보였다.

의원에서의 의사의 수 : 4명

의원에서의 환자의 수 : 7,800명

장기간 정신병 환자의 수 : 17명

표로부터 어떠한 배정이 불균형을 감소시킬지 알 수 있다. 두 번째 의원은 아래 표의 강조된 행에 영향을 미칠 것이다.

	시험군	대조군
3, 4 명의 의사	1	0
5, 6 명의 의사	0	0
8,600 명 미만의 환자	1	0
8,600 명 이상의 환자	0	0
25명 미만의 정신병 환자	1	0
25명 이상의 정신병 환자	0	0
불균형	3	0

불균형은 강조된 행의 총합이므로, 시험군에 의원 2를 배정하면 불균형은 6과 0이나, 대조군에 배정하면 3과 3이 된다. 따라서 두 번째 의원은 대조군에 배정된다.

이제 그룹간은 완벽히 균형을 이루게 되고 그러므로 3번째 의원도 무작위로 시험군에 배정되었다.

3번째 의원은 다음과 같은 특성을 보였다.

의원에서의 의사의 수 : 5명

의원에서의 환자의 수 : 10,000명

장기간 정신병 환자의 수 : 24명

이 의원은 시험군에 배정되었으므로, 할당은 다음과 같다.

	시험군	대조군
3, 4 명의 의사	1	1
5, 6 명의 의사	1	0
8,600 명 미만의 환자	1	1
8,600 명 이상의 환자	1	0
25명 미만의 정신병 환자	2	1
25명 이상의 정신병 환자	0	0

4번째 의원은 다음과 같은 특성을 보였다.

의원에서의 의사의 수 : 3명

의원에서의 환자의 수 : 3,400명

장기간 정신병 환자의 수 : 12명

이는 아래 표의 강조된 행의 불균형에 영향을 미칠 것이다.

	시험군	대조군
<b>3, 4 명의 의사</b>	<b>1</b>	<b>1</b>
5, 6 명의 의사	0	0
<b>8,600 명 미만의 환자</b>	<b>1</b>	<b>1</b>
8,600 명 이상의 환자	0	0
<b>25명 미만의 정신병 환자</b>	<b>2</b>	<b>1</b>
25명 이상의 정신병 환자	0	0
불균형	4	3

만약 네 번째 환자가 시험군에 배정된다면 불균형 총합이 7, 3이고 대조군에 배정된다면 불균형 총합이 4, 6이므로 4번째 의원은 대조군에 배정된다. 이러한 절차는 16개의 의원 모두 배정될 때까지 진행된다. 만약 다음 의원이 배정될 당시 두 그룹에서의 불균형이

같다면 무작위로 배정한다.

	시험군	대조군
3, 4 명의 의사	5	5
5, 6 명의 의사	3	3
8,600 명 미만의 환자	4	4
8,600 명 이상의 환자	4	4
25명 미만의 정신병 환자	4	4
25명 이상의 정신병 환자	4	4

두 그룹은 3변수 모두에 대해 균형을 이루었다.

이러한 최소화법은 “확률적”이지 않다는 점에 대해 거부감이 들 수도 있다. 피험자가 받게 될 치료를 연구자가 알 수 있게 되므로, 피험자의 특성은 환자를 연구에 등록시키는 연구자의 결정에 영향을 줄 수 있다. 이러한 단점을 피하기 위해서 확률화의 요소를 최소화법에 도입할 수 있다. 최소화법은 피험자가 어느 쪽으로 배정되는지 방향을 결정하기 위해 사용하지만, 실제 치료를 선택하기 위해서는 확률화법을 사용한다. 예를 들어서, 불균형을 줄이는 배정군에 배정될 확률을 2/3 혹은 3/4로 상대적으로 크게 하고, 불균형을 증가시키는 방향으로 1/3 혹은 1/4의 확률로 배정되게 한다. 보다 자세한 사항은 Pocock(1983)을 참조하라.

최소화법은 규모가 작고 다양한 그룹을 무작위배정시킬 때 가장 유용하다. 큰 규모의 연구는 일반적인 무작위배정 혹은 층화 후에 균형을 유지할 수 있다. 최소화법 혹은 층화에 사용된 변수는 그룹간의 변동을 줄일 것이므로 다중회귀와 같은 방법으로 가능한 한 분석에서 고려되어야 할 것이다(E-6.1 참조).

최소화법을 구현하는 컴퓨터 프로그램이 있으므로 시험에 등록되는 개개인을 수동으로 최소화할 필요는 없다. 또한 임상시험관련 서비스를 통해 전화나 웹을 이용한 최소화법을 제공받을 수 있다.

## B-5.9 군집

개개인이 아닌 피험자 집단 전체를 한 치료군에 할당하는 경우가 있다. 예를 들어, 건강증진프로그램 연구가 의원에서 시행되면, 치료를 의원 전체의 환자에게 적용하게 되는

것이 자연스럽고 이를테면 임상시험에 대한 광고나 정보가 환자대기실에서 제공될 수 있다. 다른 예로는 환자집단을 전염으로부터 격리해야 하는 경우가 있다. 만약 특정 간호사가 한 병동의 환자들을 돌본다면 그 간호사가 그 병동의 일부 환자들만을 돌보고 다른 환자들은 돌보지 않는 것은 어려운 일이다. 또한, 환자나 보호자에게 교육을 제공하는 경우, 교육을 받은 사람들이 대조군의 역할을 하는 다른 환자 또는 보호자에게 그들이 배운 것을 전달하는 것은 일반적으로는 바람직한 현상이나, 시험에서는 그렇지 못하다. 또 다른 예로 서비스 제공자에게 임상지침서(clinical guidelines)와 같은 치료를 제공하고, 그들의 환자로부터 자료를 수집함으로써 임상지침서의 영향을 평가하는 경우도 있을 수 있다.

한 치료에 함께 배정되는 피험자 집단을 군집(cluster)이라 한다. 군집은 연구를 설계할 때 고려되어야 하고(Kerry & Bland(1998b), Kerry & Bland(1998d), Kerry & Bland(1998e), Bland (2000)) 분석에서도 고려되어야 한다(Altman & Bland(1997), Bland & Kerry(1997), Kerry & Bland(1998), Kerry & Bland(1998c)). 군집을 사용하게 되면 시험에서의 검정력(power)이 줄어들고 연구대상수(sample size)가 더 많이 필요하므로, 연구계획서에서는 이러한 사항들을 어떻게 고려할 것인지 언급하여야 한다.

## B-5.10 시험설계

### B-5.10a 병행 그룹

임상시험에서 가장 단순한 설계는 두 그룹의 환자들끼리 병행하게 연구되는 병행그룹설계(parallel group design)이고, 이는 가장 일반적인 설계이다.

### B-5.10b 교차설계

교차시험(crossover trial)은 환자가 자신의 대조군이 되는 연구이다. 즉, 각각의 피험자가 연달아 두 (혹은 모든) 치료를 모두 받는다. 치료가 주어지는 순서는 무작위적이어야 바람직하다(B-5.5). 이 설계의 큰 장점은 치료법을 다른 피험자간에 비교하지 않고 같은 피험자 안에서 비교할 수 있다는 점이다. 이는 특히 피험자간 변동이 큰 결과의 평가에 유용할 수 있다. 교차시험은 치료의 효과가 빠른 시간 내에 나타나는 경우 가장 적절하다. 한 시기로부터 다음 시기로 넘어갈 때 치료의 잔류효과(carry-over of treatment effect)가 있을 수 있기 때문에 두 치료간에 휴약기간(wash out)이 필요할 수 있다. 만약 교차설계가 제안된다면 휴약기간에 대해 논의되어야 하고 기간을 구체적으로 명시하고 그 이유를 설명해야 한다.

### B-5.10c 그룹 내 비교

교차시험(crossover trial)은 환자 내 설계(within-patient design)이다. 환자 내 설계의 다른 형태는 같은 피험자에게 두 치료가 동시에 조사되는 것이다. 이는 치료가 눈, 귀, 팔, 다리와 같이 신체의 대응되는 부분에 독립적으로 주어지는 경우에 사용될 수 있다. 이와

유사한 설계로써 나이, 성별과 같은 요인으로 피험자들을 쌍을 맺고, 각 쌍 내에서 무작위로 치료를 배정하는 짝지은 설계(matched pairs design)가 있다. 알려진 중요한 예후인자가 있는 경우, 이 설계는 피험자간 변동의 많은 부분을 제거하고 각각의 치료를 받는 피험자들이 유사한 특성을 가지도록 한다.

#### B-5.10d 축차 설계

축차 설계(sequential design)에서는 병행 그룹(parallel group)이 연구의 대상이 되고, 한 그룹에서 충분한 이익이 있거나 두 그룹의 차이가 나타나지 않을 것이 분명할 때까지 시험이 계속된다. 피험자 각각의 자료가 수집되면 분석이 시행되므로, 결과가 아주 빠른 시간에 나타나는 경우에만 적절한 방법이다. 치료간에 큰 차이가 있으면 동등한 병행시험보다 축차 설계를 통하여 시험기간이 짧아질 수 있다. 윤리적 장점으로서는 한 치료법이 확실히 우월하거나 열등하다고 평가되면 연구를 중지하게 된다는 점이다. 병행 그룹 설계를 순차적으로 분석하는 것은 계획되지 않은 다중검정(multiple testing, E-7 참조)을 시행하게 되어 잘못된 유의한 결과를 얻을 수 있으므로 부정확한 방법이다. 순차적 분석은 연구대상수 산출과정에서 다중검정을 감안하는 것과 같이 설계단계에서 면밀히 계획되어야 한다(E-7.1, E-7.4f 참조). 보다 자세한 사항은 Whitehead(1997)를 참조하라.

#### B-5.10e 요인 설계

요인 설계(factorial design)는 다양한 요인들을 동시에 비교하는 것으로, 피험자 각각은 모든 요인들의 조합을 받는다. 예를 들어, 정맥주사 전 통증완화 연구인 EMLA 시험(Nott MR Peacock JL(1990))에서는 4개의 치료가 사용되었다(정맥주사 5분전, 10분전 EMLA, 위약, 무). 타 관심요인은 주사바늘의 크기(3종류)와 성별이었다. 이 연구는 각 치료군, 주사바늘크기, 성별의 조합에 같은 수의 피험자가 등록되어 균형을 이루도록 설계되었다. 이 설계는 연구자로 하여금 통증에 대한 치료, 주사의 크기, 성별 및 교호작용(interaction)에 대한 영향을 조사할 수 있도록 하였다. 균형 잡힌 설계가 통계분석이 용이한 것은 사실이지만, 근래 우수한 컴퓨터 프로그램의 도움으로 불균형한 설계도 큰 문제가 되지는 않는다.

요인설계는 특히 요인들의 교호작용 조사에 적절하나, 때로는 연구 가능한 피험자가 수가 제한되어 있고 요인효과가 산술적 혹은 대수적 척도로 가볍적일 때(즉, 요인 교호작용이 없을 때) 통계적 검정력(power)을 최적화하기 위해 제안될 수 있다. 계획서에 정당한 이유가 기술되지 않은 채 교호작용이 없음을 가정하는 것은 매우 강한 가정이 되므로, 계획서 심사자는 교호작용을 확인할 수 있는 연구대상수 산출방법을 제시하였는지 확인하려 할 것이다. 요인설계에서 실제로 교호작용이 없는 경우에 주요인 효과는 다른 요인들의 모든 수준 조합에 대한 평균으로 산출할 수 있으나 만약 요인효과 간에 교호작용이 있는데 이를 고려하지 않는다면, 주요인 효과에 대한 추정치는 유용하지 않을 수 있다.

## B-6 결과변수



시험 결과를 결정하는데 사용할 주결과 변수(main outcome variable)를 명확히 하는 것은 매우 중요하다. 주결과 변수는 일차결과변수(primary outcome) 또는 일차종료점(primary endpoint)이라고 한다. 하나의 일차결과변수를 선택하여 결정하는 것이 매우 중요하다. 만약 주결과 변수가 통계적으로 유의하다면 치료군이 비교군보다 더 효과가 있다고 결론 내릴 수 있다.

임상시험에서는 잠재적 부작용 등과 관련되어 있을 것으로 추정되는 부가변수(additional variable)들에 대하여도 조사하게 된다. 이런 경우에도 이차결과변수(secondary outcome) 또는 이차종료점(secondary endpoint)을 사전에 명확히 하여야 한다. 이차결과변수에 대한 통계적 분석도 수행되지만, 이 경우 주결과 변수 분석결과에 대한 해석과는 차이가 있다. 많은 이차 결과변수 각각에 대해 시행된 검정 결과 중 유의한 것이 과대해석 되지 않도록 다중 검정을 고려하여야 한다. 더욱이 이차결과변수의 통계적 유의성만으로는 치료의 효능을 평가하지 않는 것이 관례이다. 이차결과변수에서의 통계적 유의성은 결론을 도출하기 위한 근거보다는 효과를 설명하기 위한 것으로 사용되어야 한다.

시험의 주결과 분석은 주결과 변수와 이차결과변수와 관련된 연구 주제에 대한 것이어야 한다. 때로는 연구자가 다른 가설을 검정하고 싶어 할 때도 있는데 이러한 결과는 주결과 보다 강조되지 않게 주의해서 제시되어야 한다. 특별히 치료의 효과가 높게 나타난 하위그룹(subgroup)에 대한 분석 결과를 제시할 때는 더욱 신중하여야 한다. 만약 이 분석이 원래의 연구계획서에 제시되어 있다면, 하위그룹에 대한 분석결과를 제시하여 해석하는 것이 문제되지 않는다. 그러나 이러한 분석결과를 제시하는 것은 다른 하위그룹에 비하여 특정 하위그룹에서 치료효과가 좋은 것처럼 과장되어 보일 수도 있다. 또한 테스트된 다른 하위그룹에 대한 언급이 없다면, 결과가 오인될 수 있다. 만약 하위그룹에 대한 효과를 진정으로 확인하고 싶다면 다중요인분석((multifactorial analysis (E-1.1))을 실시하는 것이 권장되며 상호작용을 보정하여야 한다(A-1.6a).

## B-7 데이터 모니터링

### B-7.1 데이터 모니터링 위원회

임상시험 수행에 있어서 데이터 모니터링 위원회(data monitoring committee)가 구성되어야 하는 경우가 많다. 이 위원회는 독립적인 전문가 집단으로 임상시험 기간 동안 사전에 계획된 간격으로 소집되며, 임상시험의 진행과 수행을 감독한다. 위원회는 한 치료법이 다른 치료법보다 명백히 우월한가를 검정하기 위하여 치료군 간의 결과를 비교하기도 하고, 유해사태(adverse event)가 너무 과도하게 나타나는 게 아닌지 확인하기도 한다. 데이터 모니터링 위원회는 치료군에 대하여 눈가림 된 상태로 데이터를 중간 분석한다면, 치료에 대한 사전정보 없이 중간분석의 결과로 임상시험의 조기종료여부를 결정하게 된

다. 그러므로 위원회에 대한 눈가림은 시험의 객관적인 평가에 있어 중요하다.

## B-7.2 임상시험 조기종료 시기

우위성의 증거가 확고한 경우 시험의 조기종료가 가능하다. 자료의 모니터링은 시험의 주 분석 전에 한번이상 시행이 되므로, 통계적 유의성은 일상적으로 쓰이는 유의수준인 0.05보다 증가하게 된다. 실제로 다중검정이 행해지면 유의한 차이가 없는데도 유의한 차이가 있다고 나타나는 오류(1종 오류)가 증가하게 되므로, 유의성을 보이기 위한 임계값은 이러한 오류로 인한 시험의 조기종료가 이루어지지 않도록 수정되어야 한다. 따라서 데이터 모니터링에서는 0.001과 같이 작은 임계값을 적용하여 전체적인 유의수준이 일상적으로 쓰이는 유의수준인 0.05로 유지될 수 있도록 하여야 한다. 다중검정 시 고려하여야 할 임계값의 설정에 대해서는 Pocock(1983) and Whitehead(1997)에 보다 자세히 나와 있다. 계획된 연구대상수보다 적은 수로부터 나온 추정치는 정확도가 떨어지기 때문에, 아주 크고 유의한 차이가 관찰되지 않는 한 시험의 조기종료는 피해야 한다. 즉, 중간분석으로부터 나온 차이는 아주 커서 과학적 검토를 담당하는 위원회가 납득할 수 있어야 한다. 과거에는 임상시험이 조기종료 되는 경우는 비평의 대상이 되고, 그 결과는 무시되기도 하였다.

## B-8 동의서

### B-8.1 동의

의학연구의 윤리적인 원칙은 헬싱키 선언(Declaration of Helsinki)에 근간을 두고 있다. 모든 임상 연구원은 한번쯤은 이 선언문을 읽어 보아야 하고, 이 원칙에서 어긋나지 않는지 모든 임상시험은 반드시 면밀하게 검토되어야 한다(F3.3).

연구 계획서에는 피험자들이 모집 된 방식과, 피험자가 시험에 동의하는 과정이 포함되어야 한다. 임상시험에 참여할 피험자를 모집할 때, 연구자는 피험자의 동의와 협조를 요청하여야 한다. 연구자는 반드시 잠재적 피험자에게 시험의 목적과 그들에게 일어날 수 있는 일에 대해 알려주어야 한다. 특히 연구에 참여하지 않았을 때에도 피험자에게 아무런 불이익이 없다는 것과 연구에 참여하였을 때 피험자가 받을 수 있는 잠재적 위험성이나 이익에 대해 모두 알려주어야 한다. 만약 피험자가 임상시험에 참여할 때 그 동안 받아 왔던 치료를 중지하게 되는 경우에는 이러한 사항들이 반드시 설명 되어야 한다.

정보가 구두나 비디오 또는 문서 등 다양한 형태로 피험자에게 제공될 수 있겠지만, 반드시 피험자가 가지고 있을 수 있는 인쇄물 형태의 문서로 추가적으로 제공되어야 하고, 명료하고 단순한 언어로 표현되어야 하며, 사실만을 서술하여야 한다. 이러한 피험자 설명서(information sheets)를 작성하는 것은 쉽지 않으며, 피험자 설명서는 임상시험심사 위원회에 의해 면밀히 검토될 것이다. 시험에 등록되는 사람들은 대개 자신이 중대한 질병을 가지고 있다고 듣는 등 커다란 개인적 스트레스 하에서 시험에 참여하게 되므로 인

쇄물 형태의 피험자 설명서를 제공해야 한다. 극단적인 경우 어떤 피험자들은 연구자로부터 들은 설명사항 뿐 아니라 그들이 시험에 참여하기를 요청 받았다는 것조차 기억하지 못할 수 있기 때문에 피험자 설명서를 인쇄물 형태로 제공하는 것은 중요하다. 만약 가능하다면, 피험자들이 시험에 대한 정보를 듣고 나서, 시험에 참여하기 위한 결정을 내리는 데 가족이나 다른 사람들과 토론하기 위한 충분한 시간이 주어져야 한다.

임상시험에 참여하는 사람들은 그들이 시험에 대한 정보를 제공받았고, 그것에 동의했으며, 언제든지 동의를 철회할 수 있다는 것을 이해했다는 동의서 형식에 서명하여야 한다. 동의서 형식은 피험자 설명서와는 분리될 수 있어야 하며, 피험자는 동의서를 반환할 때에도 피험자 설명서는 가지고 있을 수 있어야 한다. 피험자 역시 서명된 동의서를 가지고 있어야 하며, 시험자는 피험자들에게 자신이 그 임상시험에 동의했다는 것을 환기시켜야 한다.

### **B-8.2 응급상황**

응급상황과 같이, 잠재적 피험자들에게 임상시험의 장점과 단점 및 여러 사항들에 대하여 설명할 시간적 여유가 없는 경우가 있다. 예를 들어, 뇌졸중의 신경세포보호제(neuroprotection)에 관한 임상시험은 발병 후 가능한 빨리 조치가 취해져야 한다. 환자가 의식이 없거나, 극단적인 혼란 상태에 있는 경우에, 피험자의 친인척들과 대리인으로부터 임상시험에 대한 동의를 받아야 한다. 그러나 대리인이 정신적으로 문제가 없는 사람을 대신해서 동의할 수는 없다. 무작위배정이 수행되고 응급처치는 피험자가 동의할 수 있는 상황에 앞서서 시행될 수 있으나 이후 피험자가 시험을 거부하는 경우에는 임상시험에서 제외되어야 한다.

### **B-8.3 어린이**

잠재적 피험자가 어린이일 경우에는 그들의 부모나 법적 보호자에 의해서 동의를 받을 수 있다. 이런 경우에 부모를 위한 피험자 설명서가 있어야 하고, 어린이가 글을 읽을 수 있을 정도라면, 어린이의 나이에 적합한 피험자 설명서를 제공하여야 한다. 피험자가 그들에게 일어나고 있는 일이 무엇인지 이해할 만큼 충분한 나이라면 동의는 그들 스스로에게서 얻어야 할 것이다. 아이들은 부모의 동의와 함께 그들의 의지에 따라서 조치를 받아야 할 것이고, 그들의 의지에 반하여 피험자가 될 수 없다.

### **B-8.4 스스로 판단할 수 없거나 의사를 표명할 수 없는 피험자**

예를 들어, 잠재적 피험자가 학습장애나, 정신질환, 마비 등 정신적으로나 신체적으로 동의할 수 없는 상황이라면 법적 보호자가 동의할 수 있다. 어린이의 경우도 가능하다면 본인에게 동의를 받아야 한다. 헬싱키 선언의 24조는 "만약 해당 연구가 제시된 사람들의 건강증진을 위한 경우가 아니라면, 그리고 이 연구가 법적으로 취약하지 않은 다른 사람들에게 의해 대신 수행될 수 없는 경우가 아니라면 이러한 대상은 연구에 포함되지 말아야 한다."라고 명시하고 있다.

### B-8.5 집단무작위배정설계

집단 무작위배정설계(cluster randomized designs)를 수행할 때에는 피험자로부터 동의를 받는 것이 더욱 어렵다. 왜냐하면, 이 설계에서는 사람들이 그룹에 따라서 임의 추출되고, 이 과정은 보통 피험자의 동의 없이 진행되기 때문이다. 예를 들어, 여러 병원에서 동시에 수행되는 임상시험의 경우, 한 병원에서 선택된 모든 환자들은 같은 치료를 받도록 무작위배정된다면 환자들은 치료를 위한 동의를 하는 것이 불가능하다. 또 다른 예로, 일반의가 무작위배정되어 질병 치료를 위한 특정 가이드라인을 받게 되는 임상시험에서, 환자들은 자신의 의사가 무작위배정되거나 자신이 특정 가이드라인에 의해 치료되는 상황에 동의할 수 있는 기회가 없다. 그것은 이미 일어난 일이기 때문이다. 환자들은 단지 임상시험을 위한 데이터를 제공할 것인지의 여부에 대하여 동의할 수 있을 뿐이다. 만약 한 병원이 특정 치료를 받기로 무작위배정되었다면, 환자들은 치료받기를 동의하여 시험에 참여하거나 거부하여 참여를 하지 않을 수 있다. 만약 치료가 특정한 병원에서 시행된다면, 그들은 임상시험 참여를 권유받았을 때 그 병원에 가는 것을 거절할 수는 있다. 그러나 일반적인 병원, 직장이나 학교에서의 건강증진프로그램(health promotion interventions)과 같은 치료에서는 심지어 피험자들이 참여거부를 할 수 없는 경우도 있다. 집단 무작위배정(cluster randomization)은 이러한 이유에서 결코 윤리적이 될 수 없다고 비판받기도 한다.

### B-8.6 무작위배정 후 동의 취득하는 경우

이와 같은 설계는 새로운 치료법을 기존치료법과 비교하는 경우에 적용되는데, 연구자는 대조군에게 새로운 치료법에 대하여 알리는 것을 원치 않을 것이다. 예를 들어, 새로운 처치는 동일한 조산사가 출산과 분만, 출생 후의 모든 것을 보살피주는 것이고, 기존 처치가 행해지는 대조군은 그 시간에 당번인 사람이 아무나 산모를 보살피고, 같은 조산사가 한 번 이상 같은 산모를 돌볼 수 없는 경우라고 해보자. 연구책임자는 임신한 여성들이 한 명의 조산사가 출산에 관련된 사항을 종합 관리하는 병원이 있다는 것을 안다면, 거의 모든 여성이 기존치료법에 만족하지 않을 것이라고 생각하여 시험에 등록된 여성들은 시험군과 대조군에 무작위배정한 후 시험군에 배정된 여성들에게 이 새로운 치료법을 수락하거나 표준치료법을 선택할 수 있게 하였다(몇 명은 기존치료법을 선택했다). 그 후 모든 여성들에게 임신부를 위한 서비스(maternity services) 연구에 동의해 달라고 요청하였고, 결과적으로 연구 데이터가 수집될 수 있었다.

이것은 무작위배정 후 동의 취득하는 경우(randomized consent design) (Zelen(1979, 1990))의 예로, 여기서는 피험자들을 무작위배정한 후, 치료나 데이터 준비를 위한 동의를 얻게 된다. 그러나 무작위배정 그 자체에 대한 동의는 획득되어지지 않는다. 이런 연구방법이 임상시험에서 사용되는 경우에는 전문가들의 심도 있는 논의가 전제되어야 하며 어떤 비평가들은 이러한 무작위 동의는 윤리적이지 않다고 비판받는 경우도 있다.

무작위배정 후 동의를 취득하는 경우(randomized consent design)는 의도된 대로 분석(intention to treat) 법칙에 따라서 분석된다(E-9). 이는 유의성을 높여주는 방법이지만, 치료 효과의 추정에 비뚤림을 가져올 수 있다.

## B-9 연구계획서 위반과 비순응

어떤 환자들은 우발적 혹은 고의적으로 연구계획서를 따르지 않거나(비순응 non-compliance), 환자의 상황에 따라 임상의가 원래 배정된 치료와 다른 치료를 하게 되는 경우가 있다(protocol violation). 이런 경우 처리할 수 있는 방법은 모든 무작위배정된 환자들을 배정된 대로 유지한 채 자료를 분석하는 것이다. 이는 기존의 무작위배정이 유지되기 때문에 그룹간의 균형을 이룰 수 있고, 치료군간의 비교가능성이 보장될 것이다. 이러한 방법을 의도된 대로 분석(intention to treat analysis)이라 한다(E-9). 의도된 대로 분석법은 환자가 임상시험에서 탈락이 된 경우에는 해당되지 않는다. 이런 경우에 결측치의 효과를 평가하기 위하여 민감도 분석(sensitivity analysis)을 수행하여야 한다.

## B-10 연구대상수 달성

임상시험에서 요구되는 연구대상수는 표준적인 검정력 계산법이나 신뢰구간의 크기를 이용하여 결정되어야 한다(D 참조). 그러나 동의할만한 수의 피험자를 모집하는 것은 매우 어렵다. 많은 임상시험에 참여한 경험을 돌이켜볼 때 충분한 피험자 모집이 어려워져서 모집기간이 예상보다 훨씬 늦어지는 경우가 많았다. 또 다른 경우는 계획된 모집대상 중 일부만으로 임상시험을 영구적으로 종료하는 것인데, 이런 경우에는 검정력은 낮아지고, 결론을 내리기에 부적절한 결과가 나올 수 있다.

많은 경우 임상시험이 시작되자마자 계획당시 기대했던 바와 달리 환자수가 희박하게 된다는 것을 깨닫게 되는데 그 이유 중의 하나는 환자가 임상시험에 동의하지 않기 때문이다. 임상시험에서 피험자의 순응도가 감소하는 것은 임상시험 약물에 대한 부정적인 평판의 근거라고 할 수 있다. 그러나 많은 부분에서 피험자 모집에 실패하는 경우는 연구자가 잠재적 대상자를 분별하는 것에서 이루어진다. 이는 새로운 치료법의 안전성이나 효능에 의심의 여지가 없기 때문이거나, 새로운 치료법이 사용되어야만 한다는 관점 때문인 경우가 많다. 예를 들어, 심근경색 환자에서 nitrate 설하정에 관한 즉각설계 임상시험(sequential trial)에서 표본수 계산에서 가정된 수보다 훨씬 적은 환자가 사망하는 것을 발견하는 경우가 이에 해당될 수 있을 것이다. 이것은 임상시험에 모집된 모든 환자들이 임상시험에서 쓰이는 약물의 적응증에 비하여 예후가 좋다는 것으로 설명된다. 고위험 환자는 임상시험에 참여하지 않은 채 nitrate 설하정 처방을 받은 경우이다.

피험자 모집이 잘 되지 않는 또 다른 이유는 스태프들이 너무 바쁘기 때문이기도 하다. 만약 진료환경이 너무 많은 환자와 적은 스태프 수로 압력을 받고 있는 상태라면, 연구

프로젝트는 관심 밖으로 제외될 것이다. 다른 문제는 그 과정에서 이미 다른 연구가 진행되고 있을 경우일 것이다. 스태프는 아마도 또 다른 임상시험에 참여하거나, 이에 필요한 피험자를 모집해야 한다는 것을 염두에 두기 어려우며, 같은 환자를 두고 각기 다른 임상시험에서 환자를 모집하기 위한 경쟁을 하게 될지도 모른다. 이런 경우에는 다른 임상시험을 위탁할 수 있는 잠재적 공동연구자를 찾는 것이 중요하다,

파일럿연구(pilot studies)(A-1.9)를 통해 연구책임자가 임상시험 수행기관의 실제적으로 피험자 모집에 관련되어 있는 스태프들과 좋은 관계를 쌓아두는 것은 이러한 문제점들을 사전에 해결 하는데 매우 유용하다. 또한 연구팀에서 피험자 모집이 얼마나 잘 되고 있는지 모니터 하는 것도 좋은 아이디어가 되는데, 이 방법은 피험자 모집의 문제점들을 조기에 발견할 수 있도록 한다.

## B-11 임상시험 종료 후

최근 헬싱키 선언은 다음과 같이 말한다:

"연구가 결론지어지는 시점에서는 연구에 참여한 모든 환자에게 연구에 의하여 밝혀진 가장 좋은 예방법, 진단법 그리고 치료법이 적용되어야만 한다."(Clause 30).

물론 이것은 급성 질환의 치료와 같은 많은 임상시험에서 가능하지 않거나 적절하지 않을 수도 있다. 그러나 만성 질환의 직접적인 환자치료에 관한 임상시험인 경우에는 꼭 언급되어야만 한다.

## C. 관찰적 연구

### C-1 환자-대조군 연구

#### C-1.1 환자-대조군 연구에서 대조군의 선택

심혈관 질환의 위험요인을 구명하는 환자-대조군 연구(A-1.4 참조)를 가정해보자. 흡연력은 명백한 위험요인으로 조사할 가치가 있다. 환자-대조군 연구를 수행하기 위해서는 연구시작시점 모집단에서 어떤 표본(병동, 병원, 일반 모집단(general population) 등)을 추출할 것인지 처음에 정해야 한다. 환자군을 병원 병동에서 모집하고, 대조군을 접근이 용이한 다른 병원에서 선택하는 경우를 생각해 볼 수 있다. 그러나 후자의 경우 선택 비뚤림이 생길 수 있다. 심혈관계 질환의 잠재적 위험인자로 흡연력이 연구 변수 중 하나인 연구를 수행하고자 할 때, 심혈관 병동에서 환자군을 선택하고 흡연력이 연구 변수 중 하나라면, 폐암과 같이 흡연과 관련이 있는 질환을 가진 환자들이 있는 병동에서 대조군을 선택하는 것은 부적절하다. 적절한 대조군의 선택은 실제적 측면과 통계적 측면 모두에서 어려운 문제이다. 이러한 문제는 Breslow & Day(1980)와 의학연구방법론(1997)에서 상세

히 논의되었다. 일반적으로 대조군을 선택하는 이상적인 방법은 환자군을 추출한 일반 모집단에서 무작위 추출하는 것이나, 이것은 모집단에서 개개인의 목록이 존재한다는 것을 가정으로 하는데 이런 가정은 성립하지 않는 경우가 많다.

### C-1.2 환자-대조군 연구에서의 짝짓기

환자-대조군 연구에서 환자와 대조군이 짝지어지는 경우도 있다. 각 환자별 한 명의 대조군이 짝지어지는 1 대 1 짝짓기 또는 환자 당 여러 명의 대조군이 짝지어지는 1 대 m 짝짓기가 있다. 후자의 경우 주어진 환자군에서 검정력을 높이기 위해 사용된다(Breslow & Day(1980)). 각 환자에 대하여 한 명 또는 그 이상의 짝지어진 대조군은 짝짓기 변수들의 값이 같거나 매우 유사하게 되고, 짝짓기 변수들은 보통 나이와 성별 등을 포함한다. 대부분 2, 3개의 짝짓기 변수들이 선택되는데, 그 이상을 선택하게 되면 적당한 대조군을 선택하여 짝을 짓는 것 자체가 어려워지는 경우가 많다. 짝짓기를 하는 이유는 짝짓기 변수에서 환자군과 대조군의 차이를 없애기 위함이다. 성취하려는 목적의 정도는 짝짓기의 정확성에 따라 좌우되며, 최대한 정확하게 짝짓기가 이루어졌을 때 균형을 이룰 수 있게 된다. 파일럿연구는 짝짓기가 제대로 이루어지는지 판단을 내릴 수 있는 좋은 방법이 될 수 있다. 연구계획서에서 짝지어진 환자-대조군 연구를 기술할 때, 이를테면 환자군과 대조군이 나이로 짝지어졌다는 것만 설명하는 것은 충분하지 않다. 1년 이내 또는 5년 이내 연령에 의한 짝짓기 등과 같이 상세하게 설명해야 한다. 또한 대조군이 어떻게 선택되었는지(예를 들어, 가능한 모든 짝짓기 대상으로부터 무작위 추출), 어떤 경우에 대조군이 탈락되는지 기술해야 한다. 대조군이 탈락되는 경우 다른 사람이 대조군으로서 선택될 것인지, 얼마나 많은 대조군이 대체될 수 있는지 설명이 되어야 한다.

짝짓기의 주목적은 교란변수를 통제하는 것이다(A-1.6 참조). 하지만 짝짓기에서 나타나는 다음과 같은 문제점이 우려되는 경우에는 다른 방법으로 교란변수를 통제하는 것이 선호되기도 한다. 짝짓기의 문제점은 다음과 같다.

1. 질병/장애의 유무에 대하여 짝짓기 변수들의 효과는 평가할 수 없다. 비록 대상 질병이 짝짓기 변수와 연관이 있다 하더라도, 짝짓기 변수들을 연구대상 변수로 고려할 수 없게 된다.
2. 짝짓기를 하는 경우 이를 통계분석에서 고려해야 하는데, 이때 분석이 지나치게 복잡해진다(E-6.1, Breslow & Day(1980)).
3. 1 대 1 짝지어진 환자-대조군 연구는 짝지어진 대상이 함께 분석되어지는데, 통계분석에서 대조군에 결측이 있다면 짝지어진 환자 또한 결측으로 처리해야 한다. 유사하게 환자군에서 결측은 그와 짝지어진 대조군의 정보의 상실을 초래한다.
4. 연구나 질병의 위험 인자들 사이에서 인과관계의 일부분을 변화시키는 변수를 짝짓게 된다면 비뚤림이 생길 수 있다. 이러한 비뚤림은 과도한 짝짓기가 원인이라고 할 수 있다.

짜짓기에 대해서는 Bland & Altman(1994c)과 Breslow & Day(1980), 역학의 원리와 응용(2005), 의학연구방법론(1997)을 참고할 수 있다.

## C-2 평가 비뚤림

식이요법과 대장암 사이의 관계를 규명하는 환자-대조군 연구를 생각해보자. 식이요법을 평가하기 위해 면접조사원이 대상자가 섭취한 음식을 조사한다고 가정하자. 만약 조사자가 환자들의 대장암 유병여부에 대해 알고 있다면, 이것은 조사자에 의해 기록된 정보와 실제 정보와의 차이인 평가 비뚤림(assessment bias)을 유발할 수 있게 된다. 조사자는 암을 가진 환자들이 실제 섭취한 음식량 보다 적게 기록할 가능성이 있다. 평가 비뚤림은 평가자가 대상 질병의 유병여부에 대해 눈가림을 하게 된다면 극복할 수 있게 되며, 의식적인 혹은 무의식적인 결과의 조작을 피할 수 있게 된다(조사자가 얼굴을 맞대고 질문하는 환자-대조군 연구에서는 눈가림이 어려울 것이나). 환자-대조군 연구에서 의무기록지를 이용하여 정보를 수집하는 경우에도 환자 혹은 대조군이라는 정보를 알게 됨으로써 평가 비뚤림이 발생할 수 있다. 따라서 이러한 경우 눈가림 방법이 선호된다.

## C-3 회상 비뚤림

환자의 결과 변수와 같은 정보를 후향적으로 수집할 때 회상 비뚤림(recall bias)은 환자-대조군 연구와 단면적 연구에서 일어나는 문제이다. 예를 들어 피험자에게 과거에 있었던 질병에 대해 묻게 되는 경우가 있다. 환자의 정보를 후향적으로 수집한다면 피험자의 회상력에 의존하므로 자료의 질은 떨어질 수 있다. 게다가 피험자의 회상 능력은 현 의학적 상태에 영향을 받을 수 있고, 관찰된 연관성을 비뚤리게 할 수 있다. 만약 회상 비뚤림이 문제가 될 수 있다면, 다른 방법을 고려할 필요가 있다. 예를 들어, 눈가림 방법을 이용하여 과거 기록에서 정보를 수집할 수도 있고, 사전에 폭로 정보가 수집되었다면 전향적인 연구로 진행할 수도 있을 것이다.

## C-4 표본조사: 표본 집단의 선택

실업과 현재의 좋지 않은 건강 사이의 연관성에 대한 연구를 수행하고자 할 때, 단면적 조사를 계획하여 서울시민에서 표본을 추출할 수 있다. 연구의 목적은 발견되는 결과를 이러한 표본에서 모집단인 서울시민으로, 나아가 국가적으로 외삽하는 것이다. 그러므로 표본에서 얻어지는 것은 최소 모집단인 서울시민을 대표해야 한다. 실제로는 전체 서울시민에서 무작위 추출을 해야만 대표적인 참 표본을 얻을 수 있다. 그렇다고 할지라도 그 표본은 단지 특정 기간만 대표할 수 있으며, 심지어 얻어진 결과를 다른 기간에 같은 서울시민에게 외삽하는 것조차 힘들 수 있다. 따라서 국가로의 확장도 어렵게 된다.

때로는 우연에 의해 무작위 표본이 원하던 만큼의 대표성을 가지지 못할 수도 있다. 예



를 들어, 실업과 건강 상태 사이의 연관성을 구명하기 위한 단면적 조사에서는 경제적 빈곤을 반영할 수도 있는 도시 내 거주지역에 대한 충분한 대표성이 있다는 것을 보장하는 것이 중요할 수도 있다. 이 경우 대표성을 보장하기 위한 하나의 방법은 층화 무작위 추출을 수행하는 것이다. 층화 무작위 추출은 표본의 대표성을 확고히 하기 위해 모집단의 정보를 이용하는 것이며 모집단의 추정치에 대해 정확성을 높일 수 있다. 여기서 거주지역을 층화요인으로 사용할 수 있다. 보통 비례적인 층화 추출을 많이 수행한다. 즉, 총 연구대상수를 층화하는 계층에 비례적으로 배분하는데 이때 분율은 총 모집단에 대한 분율로써 계층 수에 의해 결정된다. 예를 들어, 서울시민의 10%가 하나의 지역단위에 살고 있다면 전체 표본수의 10%가 이 지역에서 무작위로 선택된다.

층화는 무작위 추출의 원리를 벗어나는 것은 아니다. 층화는 선택 전에 일어나는 것이며 모집단을 계층수로 나누어 각각의 층화에 무작위로 표본 추출한다. 하나 이상의 층화요인이 있을 때에도 가능하다. 예를 들어, 거주지역으로 층화를 한 후 지역에서 나이로 층화를 할 수 있다. 이 때 너무 많은 요인으로 층화를 하지 않도록 주의해야 한다. 층화 무작위 추출을 하기 위해서는 모집단이 커야 하고 모든 구성원 및 그들의 층화요인에 대한 리스트가 있어야 한다. 층화의 수가 증가하면 시간과 비용 또한 명백히 증가하게 되지만, 표본의 대표성을 더 많이 신뢰할 수 있게 되고 결과의 일반화가 가능해진다.

## C-5 일반화와 결과의 외삽

모든 의학연구는 선택된 대상의 집단에서 수행된다. 하지만 의학연구의 유용성은 개인들의 특정 집합에 대한 정보보다는 결과의 일반화에 두고 있다. 그러나 대부분의 연구는 매우 제한적인 선정기준을 갖고 있기 때문에 일반화에 어려움이 있다. 예를 들어, 대장암과 식습관의 연관성을 조사하기 위한 단면적 연구의 참여자가 사회적 수준이 낮은 지역 사람들이라면 이러한 결과를 사회적 수준이 다른 사람들에게 외삽(extrapolation)하기는 어려울 것이다. 이러한 결과의 외삽은 명확하지 않으므로 이러한 연구의 연구자들은 사회적 수준의 폭이 큰 다른 지역과 통합하는 것을 고려해야 한다. 연구에서 사회적 수준의 폭이 큰 표본을 갖는다 하더라도, 논문을 읽는 사람은 한국에서 수행한 결과를 유럽으로 외삽하기 전 연구 참여자의 인종적 특징을 주의 깊게 살펴야 한다. 관찰적 연구들은 연구의 개개 피험자들로부터 대한 어떤 정보를 발견하기보다는 위험요인과 질병/장애와의 연관성을 조사하기 위해 수행된다. 일반화와 외삽법에 대해서는 Altman과 Bland(1998), 의학연구방법론(1997)을 보는 것이 도움이 될 것이다.

## C-6 설문 조사에서 응답률의 최대화

설문 응답률은 설문에 응한 사람의 비율이다. 설문 조사, 특히 우편 조사는 낮은 응답률을 갖는 경향이 있다(30~50%는 보통). 설문에 응답한 사람과 그렇지 않은 사람의 차이 때문에 응답률이 낮은 연구의 결과는 대상 모집단의 대표성을 보여주지 못한다. 연구계획

서에 설문지 분석이 포함되어 있다면 응답률을 최대한으로 하기 위한 계획을 기술해야 한다. 우표가 붙어 있고 주소가 적혀 있는 회신 봉투를 제공하고 참여자에게 연구의 중요성을 알려주고 익명을 보장해준다면 응답률은 높아질 수 있다. 만약 설문응답자를 확인할 수 있다면 응답하지 않은 사람들에게 다시 질문지를 보내거나 전화를 하는 방법으로 응답률을 높일 수 있다. 만약 설문응답자를 확인할 수 없다면 질문지를 재발송 하였을 때 같은 응답자로부터 중복 응답을 받을 수도 있을 것이다. 질문지에서 응답률을 높이는 것에 대해서는 Edwards(2002)를 참조할 수 있다.

## D. 연구대상수 산출

### D-1 연구대상수 산출이 제시되어야 하는 경우

연구대상수 산출은 다양한 대부분의 정량적 연구(quantitative study)들에서 요구된다. 연구대상수 산출은 정성적 연구(qualitative research, 이것은 양적 연구와 같은 단순한 기술이 아니라, 내용분석과 같은 질적 방법을 의미)에서는 필요로 하지 않는다.

연구대상수 산출은 파일럿연구에서는 요구되지 않는 경우가 있다(A-1.9 참조). 연구대상수 산출과정을 제시하지 않거나 부적절한 연구대상수 산출은 연구 계획이 승인되지 않는 가장 흔한 원인 중 하나가 될 수 있다.

### D-2 연구대상수 산출의 중요성

모집단의 특성을 추정하는 연구(예, 천식 어린이의 유병률)에서 연구대상수 산출은 요구되는 정밀성(precision) 또는 신뢰성(confidence)을 갖춘 추정값을 확고히 하는데 중요한 역할을 한다. 예를 들어, 연구대상수 20명으로부터 10% 유병률은 95% 신뢰구간이 1~31%가 되는데, 이는 정밀한 정보를 제시하지 못한다. 반면, 연구대상수 400명으로부터의 유병률 10%는 95% 신뢰구간이 7~13%가 되어 이것은 충분히 정밀하다고 할 수 있다. 연구계획서 연구대상수 산출은 연구수행 후 전자의 경우에 직면하게 되는 것을 피하는데 도움을 준다.

효과를 입증하는 연구(예, 두 치료법에서의 차이 혹은 위험 요인이 존재하는 경우와 존재하지 않는 경우의 진단의 상대적 위험도의 차이)에서의 연구대상수 산출은, 만약 임상적으로나 생물학적으로 중요하게 간주될 효과가 존재한다면 그것이 발견될 확률이 크다(즉, 분석 결과가 통계적으로 유의하다.)는 것을 입증하는데 중요하다. 만약 연구대상수가 너무 작다면, 큰 차이가 관측되더라도 표본추출(sampling)에 따른 변동에 의한 것으로 의미 있는 차이가 입증되지 못할 수 있다.

### D-3 연구대상수 산출에 필요한 정보

연구대상수 산출을 위해 통계전문가에게 자문을 구하는 것은 매우 중요하다.

연구대상수 산출에 대한 방법은 Altman(1991), Bland(2000), Armitage, Berry and Matthews(2002), 의학연구방법론(1997), 실용의학통계론(1999)과 같이 몇 개의 통계 교재에 상세히 실려 있다. 연속형 자료에서는 Machin(1998)을 이용하고, 범주형 자료일 때는 Lemeshow(1996)을 참고할 수 있다. 이러한 책들은 계산을 쉽게 하기 위해 표를 제시하고 있다. 측차 설계에서의 연구대상수는 Whitehead(1997)을 참고할 수 있다.

실제 연구대상수는 컴퓨터 프로그램을 사용하여 산출할 수 있다. 웹을 통해 쉽게 찾을 수 있는 많은 프로그램들을 사용하면 평균이나 분율의 비교와 같은 비교적 간단한 연구대상수를 계산할 수 있다. 또한 보다 많은 기능을 제공하는 nQuery advisor, Elashoff(2000)와 같은 상업용 컴퓨터 패키지도 있다.

연구대상수 산출은 다음과 같은 요인에 의존하는데, 통계전문가와 임상연구자는 이러한 내용에 대해 함께 논의해야 할 것이다.

1. 연구에서의 관심 변수와 자료의 형태(자료의 형태는 A-4, A-4.1과 A-4.2 참조)
2. 목표 검정력
3. 목표 유의수준
4. 임상적으로 유의한 효과 크기
5. 연속형 결과변수의 경우 표준편차
6. 분석에서 사용된 검정(단측 또는 양측검정)
7. 연구 설계 측면
  - 단순 무작위배정 비교임상시험(RCT)
  - 집단 무작위배정 시험
  - 동등성 시험(D-7 참조)
  - 무작위배정되지 않은 개입 연구(B-5.10c 참조)
  - 관찰적 연구
  - 유병률 연구
  - 특이도와 민감도 측정 연구
  - 연구에 찍지어진 자료가 있는 경우
  - 반복 측정된 자료가 있는 경우
  - 집단들의 크기가 다른 경우
  - 계층적 자료인 경우

무작위배정되지 않은 연구들은 차이나 연관성을 찾기 위해 수정되는 경우가 많은데 일반적으로 분석과정 내에서 교란변수를 보정하기 위해서 보다 많은 연구대상수를 필요로

한다.

고려해야할 연구대상수는 전체 모집단에 비례하는 일정 부분이 아닌 절대적으로 필요한 연구대상수이어야 한다.

## D-4 통계 용어의 설명

### D-4.1 귀무가설과 대립가설

많은 통계 분석은 두 치료의 비교, 치료법이나 환자 유형과 연관되어 있다. 관심의 차이를 요약하고 있는 수치적 값을 효과(effect)라 한다. 연구 형태에 따라서는 효과가 상관계수, 교차비 또는 상대위험도로 표현된다. 귀무가설과 대립가설을 명백히 하는 것은 연구계획에서 중요한 과정이다. 보통 귀무가설은 '효과의 차이가 없다'(차이가 0이다 상대위험도는 1이다 상관계수는 0이다)이며, 대립가설은 '효과의 차이가 있다'이다.

### D-4.2 확률값(p-값, p-value)

p-value는 귀무가설인 '효과의 차이가 없다'가 실제로 참일 때 효과의 차이가 관측될 확률이다. 보통 분율의 형태로 표현된다(예,  $p=0.03$ ).

### D-4.3 유의수준

유의수준은 귀무가설을 기각하고 효과가 있다고 결론 내릴 수 있는 p-value의 임계점(cut-off point)이다. 유의수준은 보통 5%로 설정한다( $p=5\%$ 와  $p=0.05$ 와는 같다). 만약 관측된 p값이 5%보다 작다면 실제 차이가 없을 확률은 매우 작아지기 때문에 차이가 있다고 결론 내릴 수 있게 된다.

유의수준 5%는 실제 효과가 없을 때 효과가 있다고 결론 내릴 수 있는 확률이 5%라는 의미도 된다. 따라서 효과가 없을 때 효과가 있다고 결론 내리는 것을 피하는 것이 매우 중요하다면 유의수준 1%등 보다 작은 값이 적절하다.

### D-4.4 검정력

검정력은 귀무가설이 올바르게 기각될 확률을 의미한다. 즉, 실제 효과가 있거나 연관이 있을 때 효과가 없다는 귀무가설을 기각하는 것을 의미한다. 검정력은  $100 - \text{실제 효과를 발견하지 못할 확률}$ 로 생각할 수 있다. 따라서 높은 검정력은 실제로 존재하는 효과를 발견하지 못할 확률이 작음을 의미한다. 검정력은 보통 80%, 90%, 또는 95%로 설정한다. 만약 연구에서 실제로 존재하는 효과를 입증하는 것이 중요하다면 검정력을 90%나 그 이상을 적용하는 것이 적절하다.

### D-4.5 임상적으로 유의한 효과 크기

임상적으로 혹은 생물학적으로 중요하게 고려될 수 있는 평균이나 분율의 차이 또는

교차비/상대위험도의 최소 크기이다. 적절한 연구대상수는 이러한 차이가 존재하는 경우 통계적으로 유의한 결과를 얻을 수 있게 해준다.

#### D-4.6 유의성의 단측검정과 양측검정

양측검정에서는 귀무가설이 '효과가 없다'이고 대립가설은 '어느 방향이든 효과가 있다'는 의미이다. 단측검정에서는 대립가설이 효과의 차이가 한 방향으로만 있는 것이며(예로, 치료약이 위약보다 좋다), 귀무가설은 효과가 없다 혹은 위약이 치료약 보다 좋다 중 하나로 결론 내릴 수 있다.

단측검정을 사용할 특별한 이유가 없다면 양측검정을 사용해야만 한다. 차이가 특정 방향으로 기대된다는 것만으로는 단측검정에 대한 충분한 이유로 부족하다. 임상연구자들은 때로 실제 결과가 기대했던 것과 반대 방향으로 나타나는 것을 관찰하고 놀라게 될 때도 있으며, 참 효과가 기대했던 것과 반대였다면 이는 단순히 효과가 없다는 것과는 의미가 다르다. 이러한 사항은 반드시 제대로 보고되어야 하는데, 단측검정만으로는 이를 입증하기가 어렵다. 단측검정이 적절한 경우에 대해서 살펴보려면 Bland & Altman(1994)을 참조하면 된다.

#### D-5 연구대상수 산출 시 포함되어야 할 변수

연구대상수 산출은 연구의 일차 결과변수를 이용하여 산출하여야 한다. 만약 연구가 이차 결과변수를 갖고 그것 또한 중요하다면, 연구대상수는 이러한 변수 분석을 위해서도 충분히 커야 한다. 연구대상수를 중요하다고 생각되는 변수에 따라 각각 분리해서 산출하는 것이 이 경우에 이상적이다.

#### D-6 반응률, 탈락률 등에 따른 연구대상수의 손실

연구대상수 산출은 연구종료시점까지 참여하는 연구대상수와 관련이 있다. 그러므로 기대되는 반응률과 추적손실, 순응도에 따른 대상수 손실을 가져올 수 있는 상황을 고려하여 초기 대상수를 늘려야 한다(B-10). 초기 대상수와 연구 종료 시까지 남아있는 대상수 간의 연계성은 명백히 제시되어야 한다.

#### D-7 연구 목적과 통계적 분석 방법에 따른 연구대상수 산출

대상수의 타당성은 연구의 목적에 따라 평가된다. 예를 들어 새로운 약이 기존의 약 보다 우월하다는 것을 밝히는 게 목적이라면, 연구대상수는 두 치료약의 임상적 차이를 발견하기에 충분한 수가 되어야 한다. 하지만 때때로 두 약이 동등한 효과를 갖는다는 것을 보이는 것이 시험의 목적일 수 있다. 이러한 연구는 동등성 시험 혹은 음성적(negative) 시험이라 불린다. Pocock(1983, p129-130)은 이러한 연구에서 고려할 사항에 대해 기술하

고 있으며, 이 때 요구되는 연구대상수는 차이를 입증하는 시험보다 커야 한다고 설명하고 있다. 연구계획서 검토자는 연구대상수 산출이 연구에서 기술된 목표와 연관이 있는지, 그리고 일차 결과변수에 근거하여 산출되었는지 확인해야 한다(D-5 참조).

연구대상수 산출은 연구에서 제안된 통계적 방법과도 일치해야하는데, 연구대상수와 분석방법 모두 연구의 설계에 의존하기 때문이다(E 참조). 따라서 연구대상수 산출과정에서 사용된 분석방법과 연구계획서에 제안된 통계분석 방법의 일치를 확인해야 한다.

## D-8 연구대상수 산출 및 기술에 대한 예제

단일 분율을 추정하는 연구 또는 두개의 평균이나 분율을 비교하는 연구의 연구대상수 산출은 비교적 수월하기 때문에 여기에 예를 통해 제시하였다. 하지만 연구대상수 산출은 반드시 통계전문가와 논의를 거치는 것이 바람직하다.

### D-8.1 단일 분율의 추정

아래에 제시된 식은 '정규분포에 근사한 방법'이므로 대상수가 크지 않으며 퍼센트가 0이나 100에 가깝게 추정되는 경우에 적용해서는 안된다. 이러한 경우는 '정확 방법(exact method)'을 사용하여야 한다. 이것은 일반적으로 100%에 가까운 값이 예상되는 새로운 검사법의 특이도 또는 민감도를 추정하는 연구에 해당되는 경우가 많은데 이러한 경우는 통계전문가에게 자문을 구하거나 적절한 컴퓨터 패키지를 이용해야 한다.

천식환자들 사이에서의 호흡장애 유병률에 대해 우편 설문을 실시하였다(Thomas(2001)).

요구되는 정보는 다음과 같다.

-일차 결과변수=호흡곤란 유/무

-기대되는 유병률의 최적 예측(best guess)=30% (0.30)

-요구되는 95% 신뢰구간 폭=10% (즉,  $\pm 5\%$  또는 25%에서 35%)

단일 분율 추정에 관한 연구대상수 공식은 다음과 같다.

$$n = \frac{15.4 \times p \times (1-p)}{W^2}$$

여기서  $n$ =요구되는 연구대상수,  $p$ =기대되는 분율(여기서는 0.30),  $W$ =신뢰구간의 폭(여기서는 0.10)

이러한 정보를 이용하여 산출하면,

$$n=15.4 \times 0.30 \times (0.70)/0.102=324$$

연구대상수 산출 과정에 대한 기술은 다음과 같이 쓰는 것이 좋다.

"유병률 30%에서  $\pm 5\%$ 의 95% 신뢰구간을 구하기 위해서는 천식을 갖는 총 324명의 피험자가 필요하다. 설문지에서 70%의 응답률을 고려한다면 총 480개의 설문지가 제공될 것이다."

## D-8.2 두 개의 분율 비교

다음 산출과정은 두 개 군의 분율을 비교할 때에 적용 가능하다.

목적이 미숙아에서의 패혈증 감소를 위한 군락자극인자(colony stimulating factors, CSFs)의 효능을 평가하는 위약-대조군 무작위배정 시험에서, 태어난 지 2주 정도 지난 유아에서 50% 정도 패혈증을 보인 이전연구 결과를 34%로 줄이는 것을 임상적으로 의미가 있다고 가정하였다.

요구되는 정보는 다음과 같다.

- 일차 결과변수= 출생 후 최대 72시간 안에 치료가 행해진 14일 이후에 패혈증 유/무
- 임상적으로 유의한 효과 크기=16% 또는 0.16 (즉, 50%-34%)
- 유의수준=5%
- 검정력=80%
- 검정 종류=양측 검정

2개의 분율을 비교하는 연구대상수 산출 공식(양측검정)은 다음과 같다.

$$n = \frac{[A + B]^2 \times [(p1 \times (1 - p1)) + (p2 \times (1 - p2))]}{(p1 - p2)^2}$$

여기서 n=각각의 군에 요구되는 연구대상수(2n=총 수)

p1=첫 번째 분율(여기서는 0.50)

p2=두 번째 분율(여기서는 0.34)

p1-p2= 임상적으로 유의한 효과 크기(여기서는 0.16)

A: 유의수준에 의존(표 참조)-여기서는 1.96

B: 검정력에 의존(표 참조)-여기서는 0.84

A와 B에 대한 표

유의수준	A	검정력	B
5%	1.96	80%	0.84
1%	2.58	90%	1.28
		95%	1.64

위의 정보를 공식에 대입하면,

$$n = \frac{[1.96 + 0.84]^2 \times [(0.50 \times 0.50) + (0.34 \times 0.66)]}{(0.16)^2} = 146$$

146은 각 군에 필요한 수이므로 총 수는 2배인 292이다.

연구대상수 산출에 대한 기술은 다음과 같이 하는 것이 좋다.

" 총 292명의 유아들(치료군과 위약군 각 군당 146)은 검정력 80%, 유의수준 5%로 치료 후 14일에서 패혈증 분율에 대한 군 사이의 16% 차이를 입증하기 위해 충분한 수이다. 16% 차이란, 위약군에서 50%의 패혈증 분율과 치료군에서의 34%의 패혈증 분율을 의미한다."

### D-8.3 두 개의 평균 비교

다음의 공식은 두 군의 평균을 비교할 때 적용 가능하다.

무작위배정 시험에서 자살을 시도하여 병원을 찾은 환자들에서 자살 충동 감소를 위해 일반적으로 사용되는 심리적 중재 역할을 하는 치료제를 평가하려 한다. 자살 충동은 Beck scale로 측정되었다. 이전 연구에서 이 scale의 표준편차는 7.7이었고, 5점 이상 차이가 날 때 임상적으로 의미가 있다고 한다. 중도탈락은 참여한 인원의 3분의 1 정도로 예측한다(Guthrie(2001)).

요구되는 정보는 다음과 같다.

- 일차 결과변수=자살 충동에 대한 Beck scale
- 표준편차=7.7점



- 임상적으로 유의한 효과 크기=5점
- 유의수준=5%
- 검정력=80%
- 검정 종류=양측 검정

두 개의 평균을 비교하기 위한 연구대상수 산출식(양측검정)은 다음과 같다.

$$n = \frac{[A + B]^2 \times 2 \times SD^2}{DIFF^2}$$

여기서, n=각 군에 요구되는 연구대상수(2n=총수)

SD=일차 결과변수의 표준편차-여기서는 7.7

DIFF=임상적으로 유의한 효과 크기-여기서는 5.0

A: 유의수준에 의존(표 참조)-여기서는 1.96

B: 검정력에 의존(표 참조)-여기서는 1.28

A와 B에 대한 표

유의수준	A	검정력	B
5%	1.96	80%	0.84
1%	2.58	90%	1.28
		95%	1.64

이러한 정보를 산출식에 대입하면,

$$n = \frac{[1.96 + 0.84]^2 \times 2 \times 7.72}{5.02} = 38$$

이것은 각 군당 필요한 대상수이므로, 총 수는 이것의 2배인 76명이다.

중도탈락 1/3이 고려되기 때문에 연구대상수는 각 군당 60명으로 증가하여 총 120명이

된다.

연구대상수 산출에 대한 기술은 다음과 같이 하는 것이 좋다.

"각 군당 38명의 연구대상수는 표준편차를 7.7이라 가정하고, 검정력 80%, 유의수준 5%의 경우 자살 충동의 Beck scale에서 5점이 임상적으로 의미가 있다고 할 때 충분한 대상수이다. 중도탈락이 1/3이 예상되므로, 중도탈락을 고려하여 각 군 당 60명으로 늘려야 한다(총 수 120명)."

## D-9 부적절한 연구대상수 기술

예제 1.

"이전 연구에서 150명의 피험자를 모집하였고, 매우 유의한 결과( $p=0.014$ )를 얻었기 때문에 이와 유사한 연구대상수이면 충분하다."

이전 연구에서 유의한 결과를 얻은 것은 무작위 표본 추출로 인한 '행운'일 수 있다. 연구를 제안하기 위해서는 본문에 명확한 연구대상수 산출방법이 기술되어야 한다. 검정력, 유의수준, 일차 결과변수, 이 변수에 대한 임상적으로 유의한 효과 크기, 표준편차(만약 연속형 변수라면)와 군을 비교하는 것이라면 각 군당 연구대상수가 기술되어야 한다.

예제 2.

"연구를 뒷받침할 수 있는 사전 정보가 없기 때문에 연구대상수를 산출하지 않았다."

모든 효과는 발표된 기존 정보를 찾아서 산출한 연구대상수에 근거하여 입증되어야 하거나, 아무런 사전 정보가 없는 경우 이러한 정보를 얻기 위해 작은 사전 연구를 수행하여야 한다.

표준편차에 대한 사전 정보가 없을 때 연구대상수 산출은 매우 일반적인 조건으로 계산해 볼 수도 있다. 즉, 많은 표준편차에서 발견되어질 수 있는 차이의 크기를 얻음으로써 계산해 볼 수 있을 것이다.

연구가 매우 예비적인 성격의 파일럿연구라면, 그리고 연구의 실행 가능성을 평가하거나 전체 연구의 연구대상수 산출을 위한 정보를 수집하기 위함이라면 그 연구를 위한 연구대상수 산출은 필요하지 않다.

예제 3.

"일년에 50명가량이 병원을 찾고 그 중 10%정도 연구 참여를 원하지 않는다. 연구의 기간은 2년 정도이므로 연구대상수는 90명이 된다."

비록 대부분의 연구가 실행 가능성을 고려하여 균형을 맞춰야 하지만 연구대상수가 단지 이용 가능한 피험자에 의해 결정되어서는 안된다. 이용 가능한 환자의 수가 연구의 제한 요소일지라도 연구대상수는 명확히 제시되어야 한다. 이는 연구가 갖는 임상적으로 유의한 효과 크기를 입증하기 위한 검정력이나 원하는 검정력을 적용하여 입증될 차이를 제시하기 위해서이다.

이용 가능한 피험자의 수가 임상적으로 유의한 효과 크기를 입증하기 위한 충분한 검정력을 갖지 못할 작은 수라면, 연구의 기간을 확장하거나 다기관 공동 연구로 수행하는 것을 고려해 볼 수 있을 것이다.

## E. 통계적 방법의 기술

### E-1 서론

연구계획서의 통계분석 부분은 다음의 사항들을 염두에 두고 고려하여야 한다.

- 통계적 분석 방법이 적절하게 기술되었는가?
- 용어의 사용에 모호함은 없는가?
- 연구 자료는 적절한 척도로 측정되었는가?(A-4.1 참조)
- 이 척도는 제안한 통계분석 방법을 적용하는데 적절한 형태의 척도인가?(A-4.2 참조)
- 통계분석 방법에 따른 가정이 만족되었는가? 가정이 만족되지 않았다면 어떠한 대안이 계획되었는가?
- 통계분석 방법이 연구 설계와 자료의 구조를 적절하게 설명하고 있는가?

#### E-1.1 적절한 용어의 사용

연구자가 통계분석 방법을 기술할 때 적절한 통계적 용어를 사용하는 것은 검토자들의 이해를 돕는데 매우 유용하다. 그러나 더 중요한 것은 연구자들이 그들이 사용하는 통계 용어를 실제로 잘 이해하고 사용하는가의 여부이다. 많은 연구자들은 실제로는 다중요인 분석(multifactorial analysis)을 의미하는 상황에서 자신들이 다변량 분석(multivariate analysis)을 계획했다고 말하기도 한다. 이 두 분석방법은 전혀 다른 상황에서 전혀 다른 질문에 답하기 위해 사용되며 전혀 다른 가정을 갖는다. 또한, 다변량(multivariate)과 다중요인(multifactorial)이란 용어가 연구계획서 뿐만 아니라 여러 의학논문에서도 부적절하게 혼용되어 혼란을 가중시키고 있다. 그렇다면 이 용어들은 언제 어떻게 사용되어야 하는가? 어떠한 통계분석방법에서든 결과변수를 기준으로 변동(variation)이 모형화되고 가정이 수립되게 마련이다. 반면에 설명변수는 고정된 값을 취한다고 가정한다. 이때 오직 하나의 결과변수가 관여하는 통계적 방법을 단변량(univariate)이라 하고 여러 결과변

수가 관여하는 통계적 방법을 다변량(multivariate)이라 한다. 단변량 분석은 오직 하나의 설명변수가 고려되는 단일요인(unifactorial)과 여러 설명변수가 고려되는 다중요인(multifactorial)으로 재분류될 수 있다.

## E-1.2 통계적 방법의 구체적 기술

연구자가 연구 목적을 설명하기 위하여 자료를 어떻게 사용할 것인지 세부적인 사항들에 대하여 고심했다는 것을 검토자들에게 재확인 시켜주는 것은 매우 중요한 일이다. 그러나 많은 연구자들이 연구 설계를 기술하고 연구대상수를 산출하는 일에만 집중하고 통계분석 방법을 제대로 언급하지 않고는 한다. 예를 들어 보자. '자료는 SPSS를 이용하여 분석할 것이다', '자료는 다변량 분석방법을 통하여 분석할 것이다.', '자료는 다중요인(multifactorial) 방법을 이용하여 분석할 것이다.' 등으로 통계적 방법을 기술하는 것은 매우 부적절하다. SPSS로 분석될 수 있는 통계적 방법은 수도 없이 많으며 다변량 또는 다중요인(multifactorial)으로 기술될 수 있는 통계적 방법 역시 매우 많기 때문이다. 연구자가 의도하는 통계적 방법의 명칭은 명확하게 제시되어야 한다(예, 주성분 분석, 다중회귀분석 등등). 다중요인분석과 같은 포괄적인 용어는 매우 다양한 상황에서 적용되는 통계적 방법들을 모두 포함하기 때문에 검토자들에게 명확한 통계방법에 대한 정보를 제시하는 것이 중요하다. 예를 들어, 다중 로지스틱 회귀분석과 다중 회귀분석은 둘 다 다중요인분석으로 기술될 수 있다. 하지만 후자는 결과변수가 연속형일 때에만 사용될 수 있고 전자는 결과변수가 이분형(예, 예와 아니오 같은 두 개의 범주) 이거나 이분형 변수의 합인 경우에만 사용될 수 있다(Bland(2000)의 chapter 17과 실용의학통계론(1999) 참조).

## E-2 통계적 방법의 적절성

통계적 분석 부분을 기술할 때 연구자는 연구에서 생성된 자료의 형태를 항상 염두에 두어야 한다(A-4.1, A-4.2 참조). 예를 들어, 한 연구자가 두 독립된 집단으로부터 하나의 결과변수를 측정하였다고 가정하자. 이 결과변수는 연속형 척도인가, 순서형 척도인가, 명목형 척도인가, 이분형(binary) 척도인가? 만약 결과변수가 이분형 척도라면 두 독립된 집단을 비교하기 위하여 카이제곱 검정(chi-square test)이나 Fisher의 정확검정(Fisher's exact test)과 같은 통계적 검정 방법을 사용할 수 있을 것이다. 그러나 만약 결과변수가 연속형 척도라면 독립표본 t 검정(two sample t test) 또는 Mann-Whitney U 검정을 사용할 수 있을 것이다. 결과변수가 순서형 척도라면 경향성에 대한 카이제곱 검정(chi-square test for trend) 또는 Mann-Whitney U 검정을 사용할 수 있으며 결과변수가 명목형 척도라면 카이제곱 검정을 사용할 수 있을 것이다. 각 검정방법의 적절성 여부는 자료의 형태 뿐 아니라 각 검정방법에서 요구하는 가정에 따라 달라질 수 있을 것이다(E-3, E-4 참조), 그러나 명백한 것은 결과변수가 이분형이거나 범주형 척도일 때에 독립표본 t 검정을 사용하는 것은 적절하지 않다는 것이다(유의성 검정에 관한 추가적인 사항은 Armitage, Berry and Matthews(2002), Altman(1991), Bland(2000), 실용의학통계론(1999) 참조).

### E-2.1 순위형 점수

같은 현상에 대해 여러 측면에서 정보를 얻도록 고안된 설문지에서 일련의 질문에 대한 응답에 수치를 할당하고 이를 합산하여 점수를 구성하게 되는 경우가 있다(예, 삶의 질 측정). 이러한 경우 주로 각 질문은 '예/아니오'와 같은 응답만을 요구하는 경우가 많다. 이와 같은 점수는 구간 척도라고도 할 수 없고 순서형 척도인지에 대해서도 논쟁의 여지가 있다(A-4.1, A-4.2 참조). 질문 A에 대하여 긍정적인 대답이 나오면 전체 점수에 1을 더하고 질문 B에 대해서도 마찬가지로 처리할 때, A와 B가 서로 다른 측면을 측정하기 때문에 1점의 차이가 실제로 무엇을 의미하는지는 분명하지 않기 때문이다.

이러한 점수가 여러 문항의 질문들로 구성될 때, 점수를 마치 다변량이나 다중요인분석을 위한 연속형 변수인 것처럼 다루기도 한다(E-1.1 참조). 이와 같이 다루는 주된 이유는 특별한 대안이 없기 때문이다. 그러나 이러한 분석 결과를 해석할 때에는 항상 신중을 기해야 하며 점수를 연속형 변수로 취급함으로써 발생할 수 있는 문제를 간과해서는 안 된다. 또한, 몇 개의(예를 들어, 10개 이하) 예/아니오 형태의 질문에만 근거하여 구성된 점수를 연속형 변수로 간주하는 것은 올바른 방법이 아니다. 실제로 일부 통계학자들은 순위형 점수가 절대 연속형 변수로 다루어 저서는 안된다고 주장하기도 한다. 그러므로 이러한 경우에는 단일요인 분석을 위한 비모수적 방법이 적용되어야 할 것이다(Conover(1980) 참조).

### E-3 짝지은 표본의 비교와 독립 표본의 비교

임상연구나 관찰적 연구에서(A-1 참조) 동일한 변수에 대한 두 자료 집단이 존재할 수 있으며 이 두 자료 집단을 비교하는 것이 관심이 되는 경우가 있다. 일차적으로 고려되어야 할 사항은 이 두 자료 집단이 상호 독립적으로 다루어질 수 있는가이다. 만약 두 자료 집단이 같은 개체에게서 다른 두 시점에 측정된 측정치로 구성되어 있다면 상관성이 존재하게 된다. 예를 들어, 중년 남성들에 대한 임상시험 초기의 폐기능 측정치와 5년 추적 관찰 후의 폐기능 측정치에 상관성이 존재하는 경우라 할 수 있다. 또한, 관찰적 연구에서도 이러한 경우가 발생할 수 있는데, 질병을 가진 개체(환자군)와 갖지 않은 개체(대조군)가 있고, 질병을 갖지 않은 개체가 나이와 성별과 같은 중요한 교란요인 때문에 질병을 갖은 개체와 짝지어지는 경우가 한 예라 할 수 있다. 이것을 1-1로 짝지은 환자-대조군연구(1-1 matched case-control)라 한다. 이와 같이 상관성이 존재하는 경우의 두 자료 집단 비교를 짝지은 표본의 비교(대응쌍 내의 차이 비교)이라 한다. 반면, 무작위 시험에서 개체들이 두 집단에 무작위로 할당되거나, 관찰적 연구에서 개별적인 집단들을 비교하는 것과 같이 두 자료 집단이 독립적인 경우, 두 자료 집단의 비교를 독립표본의 비교(집단들 간의 차이 비교)라 한다.

짝지은(paired) 표본의 비교와 독립표본의 비교(unpaired comparison)에는 서로 다른 통

계적 분석 방법이 적용되어야 한다. 만약 자료가 독립적인 두 집단의 측정치를 나타내는 연속형 척도라면(A-4 참조) 독립표본 t 검정(또는 Mann-Whitney U 검정)이 사용될 수 있겠지만, 만약 자료가 같은 개체에게서 측정된 전후 측정치라면 짝지은 표본 t 검정(또는 Wilcoxon 부호 순위 검정)이 사용되어야 할 것이다. 이분형 변수의 경우, 각 검정의 적절성은 이후에 설명될 다른 가정에 따라서 달라질 수 있겠지만 독립표본과 짝지은 표본 비교에 대한 검정은 각각 카이제곱 검정(또는 Fisher의 정확검정)과 McNemar 검정이 될 것이다(E-4 참조). 유의성 검정에 대한 추가적 사항은 Armitage, Berry and Matthews(2002), Altman(1991), Bland(2000), 실용의학통계론(1999)을 참조할 수 있다.

## E-4 통계적 검정의 가정

통계적 검정을 명확히 하기 위해서는 검정방법마다 특수한 가정이 요구되며 이들 가정이 만족되지 않는 경우에는 검정 자체가 유효하지 않게 된다. 그러나 많은 경우에 이러한 사실이 간과되고 있다. 예를 들어, 독립표본 t 검정은 독립된 두 집단이 등분산을 갖고 독립적인 두 정규분포로부터 무작위로 추출되었다고 가정한다. 만약, 언급한 것처럼 자료가 독립적인 두 개체 집단으로부터 추출되었다면 독립성 가정은 만족될 것이다. 한편, 자료가 정규분포로부터 추출되었는지를 확인할 수 있는 한 가지 간단한 방법은 자료의 히스토그램(Histogram)이나 정규성 도표(Normal plot)를 그려보는 것이다. 히스토그램이 대칭형의 종모양 형태를 갖거나 정규성 도표의 점들이 직선을 나타내면 정규성이 만족되었다고 볼 수 있다. 등분산 가정을 확인하기 위해서는 각 집단의 표준편차(standard deviation)를 계산하여 비교해 볼 수 있다. 이는 다른 어떠한 방법보다도 좋은 방법이라 할 수 있다. 분산들 간의 통계적 차이를 검정하는 것은 통계적 유의성이 표본수에 의존하기 때문에 한계가 있는 방법이다(표본수가 큰 연구에서는 분산들 간의 매우 작은 차이도 통계적으로 유의한 차이라고 결론을 내릴 수 있고 반면 표본수가 작은 연구에서는 매우 큰 차이도 의미가 없다고 결론을 내릴 수 있기 때문이다). 그 밖에, 짝지은 표본 t 검정(E-3 참조)은 짝지어진 관측치들 간의 차이가 정규분포로부터 추출되었다고 가정한다.

카이제곱 검정과 McNemar 검정(E-2, E-3 참조)은 대표본 검정이고 이들 검정방법의 타당성은 표본수에 의존하게 된다. 표본수가 작은 연구에서는 Fisher의 정확검정 또는 McNemar 검정의 정확 버전(exact version)이 사용될 수 있다(Cytel의 StatXact를 포함하여 몇몇 프로그램을 통하여 수행할 수 있다).

일반적으로 생각되는 것과는 달리, 순위를 기반으로 하는 Mann-Whitney U 검정이나 Wilcoxon 부호 순위 검정(E-2, E-3 참조)과 같은 검정 방법들도 몇 가지 가정을 하고 있으며 표본수가 지나치게 작은 경우에는 사용할 수 없다(Conover(1980) 참조).

### E-4.1 자료의 변환

만약 자료의 히스토그램이 대칭적이지 않고 오른쪽으로 긴 꼬리를 갖는다면 자료의 분

포가 양의 방향으로 치우쳐졌다(positive skew)고 말할 수 있다. 만약 히스토그램이 왼쪽으로 더 긴 꼬리를 갖는다면 자료의 분포가 음의 방향으로 치우쳐졌다(negative skew)고 할 수 있다. 사용하고자 하는 통계적 방법은 자료가 정규분포로부터 추출되었음을 가정하는데 실제 자료는 양의 방향으로 치우쳐진 분포를 갖는 경우에 자료의 로그변환(logarithmic transformation)이 유용한 대안이 될 수 있다. 즉 원자료에 로그를 취하고 로그변환 된 자료를 분석하는 것이다. 로그변환 된 자료의 히스토그램은 대칭적인 종모양의 적절한 분포 형태를 제시하게 되어, 로그변환 된 자료는 정규분포로부터 추출되었다는 가정을 만족하게 될 것이다. 만약 로그변환 된 자료가 정규분포를 따르지 않는다고 나타나면 다른 형태의 변환(예, 제곱근 변환(square root transformation))을 시도해 보는 것이 좋다. 많은 경우에 정규성을 만족시켜주는 자료의 변환은 등분산 가정도 만족시켜준다(Wetherill(1981) 참조).

만약 자료가 음의 방향으로 치우쳐진 분포이고 자료의 변환이 정규성을 만족시켜 주지 못한다면 순위를 기반으로 하는 비모수적 방법을 사용하는 것이 유용한 접근법이 될 수 있다. 독립표본 t검정과 동등한 비모수적 방법으로는 Mann-Whitney U 검정이 있다 (비모수적 검정에 대한 세부사항은 Conover(1980), 실용의학통계론(1999) 참조).

## E-5 교란효과의 보정

비타민 C와 폐기능 향상 사이의 관련성을 규명하고자 하는 관찰적 연구에서 비타민 C가 폐기능 향상에 직접적인 영향을 미칠 것이라고 가정할 수는 없다. 왜냐하면 흡연과 같이 비타민 C와는 역의 관계를 갖으면서 폐기능에 직접적인 영향을 미치는 교란변수가 있을 수 있기 때문이다. 즉, 흡연이 비타민 C와 폐기능 사이에 연관성을 혼동시킬 수 있다.

교란효과는 짝짓기(matching)나(C-1.2 참조) 층화무작위배정(stratified randomization)을 통하여(B-5.7 참조) 연구 설계 단계에서 보정될 수 있고 다중요인 방법을(E-1.1 참조) 사용하여 분석단계에서 보정될 수도 있다. 통계분석 방법에서는 잠재적 교란변수 리스트와 이들 교란변수들을 어떻게 보정할 수 있는가를 언급하여야 한다. 만약 연구자가 분석단계에서 교란변수를 보정하기로 결정하였다면 교란변수에 관한 자료의 수집이 연구계획의 일부로 구성되어야 하며 세부적인 정보 수집 과정이 기술되어야 한다. 폐기능과 비타민 C의 예제에서 흡연의 효과를 보정하기 위해서는 현재 흡연자, 과거 흡연자, 비흡연자를 분류하고 흡연량과 흡연기간도 중요한 요인이 되므로 이를 고려한 변수가 필요할 것이다. 만약 연구 설계 단계에서 교란변수를 보정하기로 계획하였다면 통계 분석에 적합한 구조의 자료를 구축해야 할 것이다(E-6, E-6.1 참조). 예를 들어, 1-1로 짝지은 환자-대조군 연구에서는 환자군과 대조군을 독립적인 두 집단으로 다루는 것 보다 짝지어진 집단으로 다루는 것이 더 적절하다(E-3 참조).

## E-6 계층적(다수준) 자료

계층적 자료란 같은 병원내의 환자들, 한 가족 내의 개체들과 같이 자료가 어떤 형태의 계층을 갖는 것이다. 계층적 자료에서는 같은 집단에 있는 개체들이 다른 집단에 있는 개체들 보다 유사성이 높다는 사실을 간과해서는 안된다. 이 문제는 독립성 결여의 한 형태로 볼 수 있다. T-test와 같은 대부분의 유의성 검정은 비교가 되는 각 집단 내에서(예, 치료A 또는 치료B) 자료가 이론적인 어떤 분포로부터 추출된 독립적인 관측치임을 가정한다. 예를 들어, 독립표본 t 검정은 각 집단 내의 자료가 정규분포로부터 추출된 독립적인 관측치임을 가정한다. 만약, 여러 병원으로부터 선정된 환자들에게서 총 콜레스테롤 수치를 측정한 자료가 있다고 가정해 보자. 같은 병원에서 선정된 환자들 간의 측정치는 다른 병원에서 선정된 환자들의 측정치보다 더 유사한 경향이 있을 수 있으며 그렇기 때문에 독립성 가정이 어렵게 된다.

### E-6.1 계층적 자료의 분석

정밀하게 고안된 연구에서는 자료에 계층적 구조를 도입함으로써 균형을 맞추고자 하는 경우가 있다. 예를 들어, 환자-대조군 연구에서 같은 연령과 성별의 대조군과 환자군을 1:1로 짝지을 수 있다(C-1.2 참조). 이러한 짝짓기는 연구계획서의 통계적 분석이 기술된 부분에서 설명되어야 한다(Breslow & Day(1980) 참조). 임상시험에서 연구자는 피험자를 연령에 따라 층화할 수 있고 층 내에서 무작위배정을 할 수 있다(B-5.7). 만약 정밀도(precision)를 최대화하고자 한다면 층은 통계적 분석에서 보정되어야 한다. 집단 무작위배정 시험에서(B-5.9 참조) 여러 병원으로부터 선정된 환자 표본에 대하여 두 치료법 중 하나를 무작위로 할당할 수 있고 환자들에게서 결과를 측정할 수 있다. 이러한 경우에는 환자가 아닌 각 병원이 분석의 단위로 사용된다. 즉 각 병원에 대해 요약한 결과를 사용하는 것이 타당하고 각 병원을 무시한 채 치료를 받은 군과 받지 않은 군에 있는 모든 환자들을 단순히 합할 수는 없다(Altman & Bland(1997), Bland & Kerry(1997), Kerry & Bland(1998), Kerry & Bland(1998c) 참조).

어떤 경우에는 한 수준에서 표본을 추출하고 또 다른 수준에서 실제 자료를 수집함으로써 계층적 자료의 구조가 발생하기도 한다. 이를테면, 노인가구 연구에서 노인이 있는 각 가정이 표본으로 추출되고 각 가구 내에 있는 노인들에게서 자료를 수집하는 예가 있을 수 있다. 여기서는 가구 내에 군집(clustering)에 대하여 적절한 보정이 이루어져야 하며 다수준 모형(multilevel modeling)(Goldstein(1995))과 같은 복잡한 통계적 기법이 필요할 수도 있다. 연구계획서 심사위원들은 연구자가 자료의 구조를 바르게 인식하고 있으며 제안한 통계분석 방법이 적절한지 판단해야 할 것이다.

## E-7 다중검정

### E-7.1 유의성 검정의 반복적 사용

아래와 같이 상황에 따라서는 유의성 검정을 여러 번 수행해야 하는 경우가 발생할 수



있다.

임상시험에서 두 개 이상의 결과척도가 존재하는 경우가 있다. 예를 들어, 미숙아 인공호흡 연구에서 연구자는 인공호흡 시간과 미숙아의 생존여부에 관심을 가질 수도 있고 각 변수의 차이(생존율의 차이와 인공호흡 시간의 차이)를 보고자 할 수도 있다(B-6, E-7.4a 참조).

관찰적 연구에서 두 개 이상의 예측변수가 존재하는 경우가 있다. 예를 들어, 병원입원에 대한 대기오염의 효과연구에서 연구자는 몇몇 오염물질의 효과와 시간대(예, 입원한 날의 오염정도, 입원 전날의 오염정도, 입원 전전날의 오염정도 등등)를 고려해야 한다. 만약 이들 중 하나라도 통계적으로 유의하다면 연구자는 전반적인 대기오염이 효과를 갖는다고 결론을 내리고자 할 것이다(E-7.4b 참조).

시간에 따라 반복적으로 측정된 자료(serial data)에서(예, 시험약이나 위약 투여 후 정해진 각 시점마다 측정된 순환 호르몬 수준의 측정치), 연구자는 각 시간대에서 두 집단의 차이를 보고자 할 수 있다(E-7.4c 참조).

세 집단 이상의 비교, 예를 들어, 한 연구에서 세 개의 다른 처리를 비교해야 하는 경우나 용량 수준이 두 가지인 시험약과 위약을 비교해야 하는 경우가 있다. 이러한 경우에 연구자는 두 집단씩 짝을 지어 각각을 비교하고자 할 것이다(E-7.4d 참조).

부분집단 내에서 연구가설을 검정해야 하는 경우, 즉, 남성군과 여성군 각각에서 또는 중증 질병군과 경증 질병군 각각에서 연구가설을 검정해야 하는 경우가 있다(B-6, E-7.4e 참조).

한 연구에서 환자가 추가적으로 모집될 때마다 차이에 대한 검정을 반복적으로 수행하는 경우도 있을 수 있다(B-5.10d, B-7.2, E-7.4f 참조).

연구의 주요 분석이 위에서 언급한 사항들 중 일부와 연관이 있다면 이러한 사항이 연구계획서에 기술되어야 하고 다중검정의 문제를 어떻게 해결할 것인지도 함께 설명되어야 한다.

## E-7.2 다중검정의 문제점

여러 번의 유의성 검정을 수행하게 되면 발생하는 문제는 위양성 결과(false positive results), 즉, 거짓으로 유의한 차이가 나타날 확률 또는 제 1 종 오류의 확률이 증가한다는 것이다. 다시 말하면, 모집단에 실제로 차이가 없다면 즉, 귀무가설이 사실이라면, 다중검정을 통하여 적어도 한번 유의한 차이를 얻게 될 확률이 0.05보다 커진다는 것이다(통계적 용어의 설명에 대해서는 D-4 참조).

귀무가설이 사실일 때 단 한번의 유의성 검정을 하게 되면, 거짓 양성의 확률, 즉, 유의한 결과가 나타날 확률이 정의에 의하면 0.05이고 진음성(true negative)의 확률, 다시 말해, 유의하지 않은 결과가 나타날 확률이 0.95가 된다. 만약 독립적인 두 번의 유의성 검정을 하게 되면 즉, 검정에서 사용된 변수들이 독립적이라면, 두 번의 유의성 검정에서 진음성의 확률, 유의하지 않은 결과가 나타날 확률이  $0.95^2=0.9025$ 가 된다. 그러므로 두 번의 유의성 검정에서 적어도 한번 거짓 양성이 나타날 확률은  $1-0.9025=0.0975$ , 즉 0.05보다 커지게 되는 것이다. 만약 k번의 독립적인 유의성 검정을 하게 되면 적어도 한번 거짓 양성이 나타날 확률은  $1-0.95^k$  이 된다.

14번의 독립적인 유의성 검정을 하는 경우, 적어도 한번 위양성이 나타날 확률은  $1-0.95^{14}=0.51$  이 된다. 즉, 거짓으로 유의한 차이가 있다고 나타날 확률이 50% 이상이 되는 것이다.

부분집단 분석에서 각 부분집단이 서로 겹치지 않는다면 검정은 독립적이다. 만약 검정이 독립적이지 않다면 적어도 한번 위양성이 나타날 확률은 정확히 측정할 수는 없으나  $1-0.95^k$  보다 작을 것이다. 그러나 한번의 위양성이 나타났을 때, 그 이상 위양성의 확률이 나타날 확률은 각 검정들이 독립일 때 보다는 크다. 이러한 상황을 이해하기 위하여 여러 변수들이 동일하다는 극단적인 상황을 가정해 보자. 위양성의 확률은 여전히 0.05이며  $1-0.95^k$  보다 작지만 한번 위양성이 발생하게 되면 모든 변수들이 다 유의하게 될 것이다. 그러므로 몇몇 유의한 결과들을 보게 됐을 때 그것이 위양성이 아니라고 믿을 수 있을지 알 수 없게 되는 것이다.

### E-7.3 Bonferroni 보정

다중검정의 문제를 해결하는 통상적인 방법 중 하나는 Bonferroni 보정을 사용하는 것이다. 모든 귀무가설이 사실이고  $\alpha$ 의 임계 p-value를 사용하여 k번의 유의성 검정을 한다고 가정해 보자. 적어도 한번 유의한 차이가 나타날 확률은  $1-(1-\alpha)^k$  이다. 이것을 우리가 원하는 유의수준 0.05라고 하자.  $\alpha$ 는 매우 작아질 것이고  $(1-\alpha)^k=1-k\alpha$  와 같은 근사적인 수치를 사용할 수 있다. 그러므로  $1-(1-\alpha)^k=1-(1-k\alpha)=0.05$  가 되고,  $k\alpha=0.05, \alpha=0.05/k$  가 된다. 따라서 만약 k번의 유의성 검정을 하고 이들 중 한번의 유의성 검정에서  $p\text{-value}<0.05/k$  을 만족하게 되면, 모든 k개의 귀무가설이 사실이라는 복합귀무가설에 대한 p-value는 0.05가 된다. 실제로 복합귀무가설의 검정이 유의수준 0.05에서 유의하다면 개별적인 p-value에 k를 곱하는 것이 좋은 대안이 될 수 있으며 k가 곱해진 수정된 p-value들 중에서 가장 작은 값이 복합귀무가설에 대한 p-value가 된다 (Bland & Altman(1995), Bland(2000a)).

Bonferroni의 보정은 모든 검정이 독립임을 가정한다. 검정이 독립이 아닐 때 Bonferroni의 보정을 적용하게 되면 원래의 값보다 p-value가 커지게 되며 그 커지는 정도가 얼마만큼인지는 계산할 수 없다. 그러므로 그 연구의 검정력이 어느 정도인지 파악할 수 없게 된다. 가능하다면, 자료의 구조를 설명할 수 있는 Bonferroni 외 다른 방법을 찾아보는 것이 현명할 것이다.

해당 연구에서 여러 번의 유의성 검정을 해야 하는 경우에 연구자는 연구계획서에서 이들을 어떻게 다룰 것인지 기술해야 한다(E-7.4 참조).

#### E-7.4 다중검정 문제의 해결

다중검정의 문제가 발생했을 때 이를 무시하고 각 검정을 수행할 수도 있을 것이다. 그렇지만 이는 분석결과를 접하는 사람들에게 오해의 소지를 불러일으킬 수 있으므로 좋은 방법이라 할 수 없다.

연구자는 하나의 주요한 검정을 선택하고 이를 고수하는 것이 중요하다. 이는 확증을 위한 임상시험에서는 좋은 방법이지만 중요한 정보를 간과할 수도 있기 때문에 다른 연구 설계에서는 실용적이지 않을 수 있다.

다중검정의 문제에서는 다중검정이 발생하게 되는 형태에 따라서 여러 가지 다른 해결 방법이 있을 수 있다(E-7.4a, E-7.4b, E-7.4c, E-7.4d, E-7.4e, E-7.4f 참조).

##### E-7.4a 하나의 임상시험에서 여러 개의 결과변수가 있는 경우

결과변수의 수를 최소화하는 것이 가장 바람직하지만 어쩔 수 없이 여러 번의 검정을 수행해야 하는 경우가 발생하게 된다. 이러한 경우에 Bonferroni의 보정방법(E-7.3 참조)을 사용할 수 있다. 즉, 관측된 p-value에 수행된 검정의 수를 곱하는 것이다. 만약 수정된 p-value 중 적어도 하나의 값이 0.05보다 작다면 처리집단 간에는 통계적으로 유의한 차이가 있다고 말할 수 있다. 이는 검정하고자 하는 변수들 중 어떤 변수에서도 처리에 따른 차이가 없다는 복합귀무가설을 검정하는 것이다. 마찬가지로 Bonferroni 보정방법은 연구대상수 산출 시에도 고려되어야 한다. 만약 "처리 간 차이가 있다" 라는 결론을 이끌어 수 있는 두, 세 개의 주요 결과변수가 있다면 이는 기존의 제 1 종 오류를 변수별 검정의 수로 나누어 주어야 한다. 또한, 측정치에 대하여 여러 개의 검정을 수행하고자 할 때에도 제 1 종 오류가 보정되어야 하며, 위의 모든 경우에 연구자는 연구계획서에서 검정법이 Bonferroni 방법으로 보정되었음을 기술하여야 한다. 또는, 이와 같은 다중검정은 향후 가설 수립을 위한 것이므로 어떠한 확증을 이끌어내기 위한 것이 아님을 분명히 밝히는 것이 좋다.

##### E-7.4b 관찰적 연구에서 여러 개의 예측변수가 있는 경우

일반적으로 다중검정의 문제를 발생시키는 변수들이 서로 관련되어 있지 않다면 이를 무시하게 된다. 이러한 상황이 바로 Bonferroni 보정법(E-7.3 참조)을 사용할 수 있는 상황이며, 예측변수들 각각을 검정하고 p-value에 대하여 보정을 해야 한다. 또한, 연구계획서에서는 0.05를 검정의 수로 나누어 제 1 종 오류로 사용하고 연구대상수 산출 공식에 적용하여야 한다. 한 가지 대안은 다중 회귀분석이나 다중 로지스틱 분석과 같이 다중요인분석에서 여러 변수들을 한꺼번에 검정하는 것이다. 이러한 경우에는 개별적인 p-value를 무시해도 되지만 이러한 분석을 안전하게 시행하기 위해서는 충분한 표본수가 확보되어야 한다.

#### E-7.4c 시간에 따라 반복측정 된 자료의 경우

각각의 시간대에서 개별적으로 검정을 하는 것은 바람직하지 못한 방법이다. 이러한 방법은 거짓 양성의 확률을 증가시킬 뿐 아니라 자료를 비효율적으로 사용하는 것으로서 거짓 음성의 확률까지도 증가시킨다. 이러한 상황에서의 가능한 접근법 중에 하나는 AUC(the area under the curve)와 같은 요약통계량을 만드는 것이다(Bland(2000b), Matthews et al.(1990)). 측정된 변수의 최대값과 최대값이 나타나는 시간을 사용할 수도 있지만 이 방법 역시 측정된 모든 자료를 사용하지 않기 때문에 비효율적인 방법이다. 그렇지만 이 방법은 해석이 용이하고 회귀직선의 기울기로 측정되는 변화율이 증가하거나 감소하는 자료의 경우 좋은 요약통계량이 될 수 있다. 연구자는 연구계획서에서 사용하고 자 하는 요약통계량이 무엇인지 결정하여야 하고, 연구대상수 산출을 위해 표준편차와 검증하고자 하는 차이의 크기에 대한 추정치가 필요하게 될 것이다. 이러한 경우에는 파일럿연구(A-1.9 참조)가 매우 유용할 수 있다. 또 다른 접근 방법으로는 반복측정자료의 분산분석법이나 다수준분석법이 있다(Goldstein(1995)). 이러한 방법은 분석을 수행하고 해석하는 면에서 훨씬 복잡한 과정이 필요하게 되므로 숙련된 통계학자가 연구에 반드시 참여하는 것이 바람직할 것이다.

#### E-7.4d 세 집단 이상을 비교하는 경우

여기서는 분산분석이나 다른 여러 집단 비교 방법을 이용하여 설명할 것이다. 만약 전체 집단들 간에 유의한 차이가 있다면 연구자는 각 집단들을 짝지어서 비교해 보고자 할 것이다. 이를 수행하기 위한 몇 가지 방법이 있다. 첫 번째 방법은 분산분석의 잔차분산(residual variance)을 이용하여 짝지어진 각 집단들 간에 t-test를 수행하는 것이다. 이를 최소유의차검정이라 한다. 그러나 다중검정으로 인하여 거짓 양성의 확률이 높아지게 된다. 물론, Bonferroni의 보정법을 사용할 수 있으나 검정력이 떨어지게 되고 복합가설(composite hypothesis)을 검정하는 경우가 아닐 때는 그리 적절하지 않다. 이보다는 귀무가설이 사실일 때 20번의 분산분석을 시행해서 단지 한번 유의한 차이가 발견되는 성질을 갖는 방법이 20번의 짝비교 중에 한번 유의한 결과를 보이게 되는 방법보다 더 좋은 방법이라 할 수 있을 것이다. 이러한 방법으로는 각 집단의 표본수가 동일한 경우에 적합한 Newman Keuls range test, Gabriel's test 등이 있으며, 통계 패키지마다 다른 접근 방법으로 이와 같은 검정을 수행한다. 따라서 이러한 방법을 사용하기 위해서는 보유

하고 있는 통계패키지가 무엇인가에 의존해야하는 제한점이 있을 것이다. 소프트웨어 별 다양한 방법들을 일일이 검토하여 어떤 소프트웨어를 사용해야 하는지를 제시할 수는 없겠으나, 연구자는 연구계획서에 사용하고자 하는 방법이 무엇인지는 명시해야 할 것이다. 세 집단 이상 비교에서 연구대상수 산출은 매우 까다로운 일임을 인식해야 할 것이다.

#### E-7.4e 부분집단 내에서 연구가설을 검정하는 경우

부분집단 내에서 연구가설을 검정하고자 하는 경우는 두 가지가 있을 수 있다. 첫째는 전체적인 처리의 차이는 유의하지 않더라도 어떤 특정한 환자 집단에 대해서는 유의한 처리의 차이가 있는지 보고자 하는 경우이다. 만약 그렇다면, 연구자는 치료가 효과적이라고 결론을 내리고 싶어 할 것이다. 이러한 상황에서는 복합귀무가설을 검정해야 하는 상황처럼 Bonferroni의 보정법을 필요로 하지 않는다. 각 검정이 독립적이기 때문에 구태여 검정력을 잃을 필요는 없다. 둘째는 주요 처리의 차이가 부분집단들마다 다양하게 나타나는 경우로서 예를 들어, 처리의 효과가 증상이 중중인 집단에서 경중인 집단보다 크게 나타나는 경우이다. 이러한 경우에는 주요 처리 요인과 부분집단 요인 간에 교호작용 효과를 추정함으로써 문제를 해결할 수 있다. 부분집단 내에서의 개별적인 검정을 통해서 이러한 상황을 제대로 설명할 수 없다. 개별적인 검정을 하게 되면 다중검정을 해야 할 뿐 아니라 하나의 검정에서는 유의한 차이가 있었고 다른 검정에서는 유의한 차이가 없었다고 해서 두 부분집단 간에 유의한 차이가 있다고 결론을 내릴 수도 없기 때문이다. 이러한 상황에서는 통계적으로 유의하지 않다는 것이 처리효과가 없음을 의미하지 않는다.

#### E-7.4f 환자가 등록됨에 따른 반복적인 검정

이것은 전통적으로 다중검정과 관련된 문제이다. 이러한 문제는 다중검정이 고려되고 이것이 연구대상수 산출에 반영되는 축차설계방법을 적용함으로써 해결할 수 있다. 이와 같은 설계에 따른 검정은 전체 p-value가 0.05를 만족할 수 있도록 보정된다. 축차설계에는 다양한 방법이 있으며 축차설계를 적용하고자 하는 연구자는 Whitehead(1997)를 참고할 것을 권한다.

### E-8 평균값으로의 회귀 현상

연구목적 중 하나가 초기값과 시간에 따른 변화 값의 연관성을 보고자 하는 것이라면 여러 가지 문제점이 발생할 수 있다. 측정오차로 인하여, 높은 초기값을 갖는 측정치들은 추적관찰 시에 더 큰 값을 얻기 보다는 작은 값을 얻게 되는 경향이 있고 낮은 초기값을 갖는 측정치들은 더 작은 값 보다는 큰 값을 얻게 되는 경향이 있다. 그러므로 초기값과 변화값에 대한 잘못된 반대 연관성이 나타날 수도 있다. 이러한 현상은 평균값으로의 회귀(regression towards the mean)의 한 예로 볼 수 있다. 연구자가 초기값과 변화값의 연관성을 측정하고자 한다면 가장 먼저 평균값으로의 회귀효과를 제거해야 한다.

만약 연구자가 시간에 따른 변화값이 초기값에 의존적이라고 생각한다면 다중요인분석에서 초기값에 대한 보정을 시도할 수 있다. 그러나 만약 연구자가 결과변수로 변화값을 주고 설명변수로 초기값을 주게 된다면 평균값으로의 회귀로 인한 비뚤림이 야기될 것이다. 이러한 비뚤림은 초기값 보다는 평균값( $(\text{초기값} + \text{변화값})/2$ )에 대한 보정을 통하여 감소시킬 수 있다. 관찰적 연구에서도 초기값에 대하여 보정하고자 하는 시도가 경마효과로 인하여 과도하게 보정될 수 있다. 경마효과란 관심이 되는 결과변수(예, 폐기능)에서 시간에 따라 빠른 감소율을 갖는 환자들이 과거의 감소율로 인하여 더 낮은 초기값을 갖는 경향을 의미한다.

## E-9 임상시험에서의 의도된 대로의 분석원칙

무작위 임상시험에서 환자들은 무작위로 특정 집단에 배정된다. 이러한 무작위배정의 목적은 처리 결과에 영향을 미칠 수 있는 처리 이외의 요인들에 있어 초기 상태에서 동질적이고 비교성 있는 표본을 만들고자 함이다. 효과면에서 이들은 동일한 모집단으로부터의 무작위 표본이라고 간주할 수 있다. 그러나 임상 시험이 진행되면서 어떠한 환자들은 그들이 받고 있는 치료를 변경하게 될 수도 있고 할당된 치료를 그만둘 수도 있다. 이렇게 되면 원래 환자들이 할당 받았던 치료 보다는 실제로 환자들에게 주어진 치료에 따라서 분석을 하는 것이 당연하게 여겨질지도 모른다. 그러나 이러한 접근은 애초에 무작위배정으로 인해 부여되었던 군간 비교성을 유지할 수 없게 한다. 또한, 치료를 변경하거나 중지한 환자들은 통상적인 경우가 아닐 수 있다. 실제로 이러한 환자들은 그들이 받고 있는 치료가 제대로 효과를 보이지 않거나 부작용이 나타난다는 이유 등으로 처음 할당 받은 치료를 변경하게 되는 경우가 있다. 그러므로 무작위 임상시험에서 의도된 대로의 분석원칙을 고수하는 일은 매우 중요하다. 이것은 통계분석에서 모든 환자들을 그들이 실제로 받은 치료에 상관없이 처음 배정 받은 군에 남겨두는 것을 의미한다. 의도된 대로의 분석원칙에 대한 기술은 무작위 임상시험의 통계분석계획 부분에서 언급되어야 할 것이다(B-9 참조).

## E-10 집단 무작위배정 시험

서로 다른 병원과 같이 환자들이 군집을 이루고 있을 때에는 이러한 사항이 분석 계획에 언급되고 고려되어야 한다(B 5-9, E-6, E-6.1 참조).

## E-11 변수의 통합

정보가 연속형 척도로 수집되었다면 통계적 분석에서는 범주화된 변수가 아닌 수집된 그대로의 연속형 변수가 사용되어야 한다. 통계분석 이전에 연속형 변수를 범주화하는 것은 연속형 변수의 정밀도(precision)가 떨어지게 기록되었거나 변수의 범주화가 가장 현실적인 접근인 경우를 제외하고는 근본적으로 이용 가능한 정보를 상실하는 것이다. 단순

히 자료를 기술하기 위한 목적으로 연속형 변수를 범주화하고자 하는 생각은 괜찮지만 유의성 검정을 적용하고자 하는 경우에는 연속형 변수를 그대로 사용하는 것이 적절하다. 예를 들어, 지방함량이 풍부한 생선의 섭취와 폐기능과의 연관성을 보고자 하는 연구에서 연구자는 생선 섭취량의 분포를 5개의 범주로 분류하여 이에 따른 폐기능을 기술할 수 있다. 그러나 지방함량이 풍부한 생선의 섭취량과 폐기능과의 연관성을 검정하고자 한다면 연속형 변수인 생선 섭취량을 그대로 사용하는 것이 바람직하다. 만약 생선의 섭취량과 폐기능과의 연관성이 직선이나 곡선의 형태를 따르지 않고 계단 형태나 비연속적인 형태라고 확신할 수 있는 강력한 이유가 있다면 예외가 될 수도 있을 것이다. 즉, 지방함량이 풍부한 생선을 섭취한 양 보다는 생선의 섭취 여부에 따라 폐기능이 영향을 받을 수 있을 것이라는 강력한 확신이 있는 경우가 이러한 예가 될 수 있다. 또 다른 예외는 생선을 섭취한 사람의 수가 너무 적은 경우이다. 이러한 경우에도 생선의 섭취는 이분형 변수로 분석될 수 있을 것이다.

## E-12 추정과 신뢰구간

두 집단 간 결과변수의 차이가 통계적으로 유의한지 찾아내고자 하는 대부분의 연구에서는 차이의 양(magnitude), 관련성의 정도가 주요한 관심이 된다. 즉, 많은 연구에서 연구 목적들 중 하나는 관심이 되는 치료가 나타내는 효과의 양, 질병의 유병률, 선형 관계의 정도, 특정 노출과 관련된 상대 위험도, 진단 검사법의 민감도 등을 추정하는 것이다. 이러한 모든 예에서 연구자는 단일 표본 혹은 단일 연구를 통하여 보다 넓은 모집단에 대한 특성을 추정하고 싶어 한다. 그러나 연구대상수에 차이가 있는 다양한 연구들은 서로 다른 추정치를 산출해 낼 수 있음을 염두에 두어야 한다. 그러므로 추정치를 제시할 때에는 연구대상수에 따라 달라질 수 있는 추정치의 변동의 척도를 함께 제시하는 것이 중요하다. 이러한 추정치의 변동의 척도를 제시하는 유용한 방법들 중 하나는 신뢰구간을 계산하여 제시하는 것이다. 95% 신뢰구간이란 동일한 100번의 연구를 수행하는 경우 95번은 연구자가 추정하고자 하는 모수값을 신뢰구간 안에 포함할 수 있음을 의미한다. 90% 신뢰구간에서는 동일한 100번의 연구를 수행하는 경우 90번 모수값을 찾아낼 수 있을 것이다. 신뢰구간은 추정치의 표본오차 값을 근간으로 하기 때문에 연구대상수에 따른 추정치의 변동을 반영하고 있다. 신뢰구간은 모수값이 취할 수 있는 최대 범위와 최소 범위에 대한 아이디어를 제공해 줄 수 있다. 또한 신뢰구간이 매우 넓다면 추정치가 정확하지 못하다는 정보를, 신뢰구간이 매우 좁다면 추정치가 매우 정확하며 유용하다는 정보를 얻을 수 있다.

대부분의 연구에서 신뢰구간의 추정은 통계분석의 중요한 부분을 차지하고 있으며 그러므로 신뢰구간의 제시 계획과 어떻게 계산할 것이라는 내용이 연구계획서의 통계분석 부분에서 언급되어야 한다. 또한, 통상적으로 95% 신뢰구간이 제시되기 때문에 연구자가 90% 신뢰구간이나 99% 신뢰구간을 제시한다면 이에 대한 적합한 이유를 언급해 주는 것이 좋다. 신뢰구간의 계산방법은 자료의 종류에 따라 달라질 수 있으며 유의성 검정의

가정에 따라 신뢰구간의 타당성이 달라지게 된다. 예를 들어, 연구자가 독립표본 t 검정 방법을 이용하여 두 집단의 평균차이에 대한 95% 신뢰구간을 추정하고자 한다면 기본적으로 자료는 연속형 척도를 가지고 있어야 한다. 또한, 두 집단은 정규분포를 하며 분산이 동일한 모집단의 표본이어야 한다(독립표본 t 검정에 대해서는 E-4 참조). 연구계획서 심사자들은 연구자가 가정에 대해 이해하고 있는지, 만약 가정이 만족되지 않았다면 무엇을 해야 하는지 알고 있는지를 확인해야 할 것이다(Altman et al.(2000) 참조).

유의성 검정보다 추정이 연구의 주요한 목적이라면 신뢰구간 역시 연구대상수 산출의 고유한 부분을 형성하게 될 것이다. 연구대상수를 산출하는 방법은 신뢰구간의 넓이에 의해 결정되는 적절한 정밀도(precision)를 가지고 추정되어질 모집단의 특성에 따라 선택될 수 있다.

### E-12.1 분율값이 1 또는 0 에 가까운 경우

민감도와 특이도 연구에서와 같이 추정하고자 하는 분율이 1 또는 0에 가까운 경우에는, 이들의 신뢰구간이 대칭적인 형태를 나타내지 않는 경향이 있다. 따라서 이러한 경우에는 대표본에 근거하여 정규분포를 가정하는 방법 보다 이항분포를 직접 적용하는 정확 검정 방법을 이용하여 신뢰구간을 추정하는 것이 바람직하다. 95% 정확 신뢰구간은 StatXact (by Cytel) 나 CIA (Altman et al.(2000))와 같이 전문적인 통계 소프트웨어를 이용하여 계산할 수 있으며 SPSS, SAS 등의 상용화된 통계 소프트웨어에서도 이용이 가능하다.

## F. 일반적 고려사항

### F-1 통계 전문가 (통계 분석)

아직 존재하지도 않는 데이터에 대해 적합한 분석 방법을 설명하는 것이 임상 연구자에게 의미 없어 보일 수도 있으며, 특히 데이터의 분포(예를 들어 정규분포일지조차 모르는 상황)나 데이터에 대한 가정조차 확실하지 않은 상황에서 정확한 접근방법을 결정하기란 어려운 일이다. 그러나 통계분석을 의뢰받은 통계전문가들은 너무나도 자주 분석조차 불가능한 데이터를 접하게 된다. 이러한 상황을 일으킬만한 연구 설계상의 문제점들은 연구계획서를 작성하는 과정에서 실제 자료가 어떻게 분석될 것인지를 고민할 수 있어야 알게 된다. 그러므로 연구계획단계에서 통계 분석 방법에 자신이 없다면 통계전문가에게 자문을 구하는 것이 가장 바람직하다.

통계 분석법에 대해 고민해보거나, 통계전문가와 상의해본 경험이 있는 연구자들이라면 통계 전문가의 필요성을 절실히 느꼈을 것이다. 통계 전문 지식은 통계전문가로부터 조언을 통하여 얻을 수도 있고, 보다 이상적으로는 통계 전문가를 공동 연구자로 추가하여 연구 전반을 함께 논의하고 수행해 나가며 얻을 수도 있다. 통계 분석 방법이 복잡할수록



연구계획서를 검토하고 심사하는 사람들은 연구 수행됨이 복잡한 통계문제를 이해하고 적절하게 해석을 할 수 있는 능력이 있는지 확인하고자 할 것이다.

## **F-2 통계 프로그램/패키지**

연구자들은 그들이 합법적으로 통계 프로그램을 보유하고 있는지, 혹은 연구비 예산에 포함시켰는지를 확인하여야 한다. 상업적인 목적으로 사용되는 여러 통계 프로그램 패키지가 있는데, 그 예로 STATA, SAS, SPSS, GENSTAT, StatXact, MLWin, S-PLUS, BMDP, MINITAB, CIA 등이 있으며, 각각의 패키지마다 다양한 특성을 가지므로 연구 목적과 자료분석방법에 따라 가장 적절하고 사용가능한 통계프로그램이 무엇인지 파악할 수 있어야 할 것이다. 불법적으로 보유하고 있는 통계패키지를 연구계획서나 차후 결과보고서에 언급하게 될 경우 문제가 야기될 수 있음을 명심하는 것도 중요하다. 웹사이트에서 다운로드가 가능한 간단하지만 제한된 기능만을 갖춘 무료 통계 프로그램들도 다수 있다.

## **F-3 윤리**

### **F-3.1 통계의 오용**

Altman은 통계의 오용이 비윤리적일 수 있음을 강력히 주장한 바 있다(1991, 1982). 통계의 오용이라는 용어는 잘못되거나 오해의 소지가 있는 결론을 유도할 수 있는 부적절한 연구 설계와 부적절한 통계 방법의 사용 모두를 포함한다. 이러한 경우 환자는 불편함을 감수하거나 불필요한 비용을 부담하는 것으로부터 시작하여 잘못된 연구 결과에서 근거한 잘못된 임상 의사결정으로 인하여 피해를 입을 수 있다.

### **F-3.2 비판적 평가**

Altman은 선행 논문들의 결과를 액면 그대로 받아들일 것이 아니라 비판적으로 검토하여야 할 필요성을 강조함으로써 윤리에 대한 논쟁을 확장시켰다(1991, 1982). 비판적으로 평가되지 않은 잘못된 결과는 향후 전적으로 불필요한 방향으로의 연구를 유도할 수 있다. 잘못된 선행 연구 결과에 대한 비판적 평가는 연구 계획서 작성 시 연구배경부분에서 언급될 수 있는 매우 적합한 내용이다. 선행 연구의 결과가 비판 없이 그대로 제시되어서는 안 된다.

### **F-3.3 사람을 대상으로 한 연구**

임상 시험 등의 사람을 대상으로 한 연구에는 많은 윤리적 쟁점이 따른다. 이러한 쟁점에 대한 지침서로는 헬싱키 선언(Declaration of Helsinki)이 있다. 윤리적 원칙에 대한 선언문은 1964년 헬싱키에서 개최된 World Medical Association 의 총회 General Assembly Meeting 에서 처음으로 차용되어 1975년, 1983년, 1989년, 1996년 그리고 2000년에 개정되었다. 의학연구를 수행하고자 하는 연구자는 모두 필수적으로 읽어야 하며,

임상시험을 수행하기를 원하는 경우는 더욱 그러하다.

많은 주류 저널들은 헬싱키 선언에 따라 연구가 수행된 경우에만 출판을 허가하고 있다. 헬싱키 선언에서 제안된 지침이 어떠한 방식으로 이행되어야 할지는 여전히 논의의 대상이지만, 최근의 지침들은 비교임상시험에 초점이 맞춰져 있다. Rothman 등은 신약이 위약이 아니라 효과가 증명된 대체 약물과 비교되어야 하고, 위약을 사용하는 경우 위약군은 비윤리적인 상황에 처하는 것이라 주장한 바 있다(2000). 2000년 스코틀랜드 회의에서 World Medical Association은 현재 principle No. 29[<http://www.wma.net> 혹은 B-3 참조]에 기술된 견해를 채택하였다. 임상시험에 관련된 윤리적 고려사항은 B-1, B-3 그리고 B-8 장에 기술되어있다.

## **F-4 기타 연구 쟁점**

### **F-4.1 연구의 관리 감독**

모든 연구는 과학적이고 윤리적인 기준에 의하여 수행되어야 한다. 연구에 대한 관리 감독은 이러한 기준들이 만족되었는가를 확인하는 과정이다. 연구의 후원자나 자금 제공자들은 임상시험 및 의학연구가 이러한 기준들을 따를 뿐만이 아니라, 그러한 과정을 확인하기 위한 절차가 마련되어 있는지 확인할 필요가 있다.

### **F-4.2 데이터의 보호/기밀유지**

사람을 대상으로 하는 모든 연구는 데이터 보호 원칙 하에서, 개인 정보의 기밀유지를 보장하고 데이터의 활용을 동의서의 범위 내로 제한하여 수행되어야 한다. 영국과 같은 경우 이에 대한 관리의 책임이 정보 관리국(Information Commissioner)의 소관으로 되어 있다. 데이터 보호 원칙과 법률을 따르는 방법 등에 관한 내용이 연구계획서에도 포함되어야 할 것이다.

## 부록 1 임상시험 계획서 작성 시 체크 리스트

해당 사항에 표시하십시오.

### 연구 설계

- 관찰적 연구인가? [ ](A-1.1, A-1.2 참조)
- 관찰적 연구라면,
  - 코호트 연구인가? [ ](A-1.3 참조)
  - 환자대조군 연구인가? [ ](A-1.4 참조)
  - 단면적 연구인가? [ ](A-1.5 참조)
  - 단면적 연구라면,
    - 유행률 연구인가? [ ](A-1.5a 참조)
    - 민감도와 특이도 추정을 위한 연구인가? [ ](A-1.5b, A-1.5c 참조)
    - 생태학적 연구인가? [ ](A-1.5d 참조)
  - 연구 계획서에 교란변수에 대한 문제점을 기술하였는가? [ ](A-1.6 참조)
- 실험적 연구인가? [ ](A-1.1, A-1.2 참조)
- 실험적 연구라면,
  - 임상시험인가? [ ](A-1.7, A-1.8 참조)
  - 임상시험이라면,
    - 대조군이 존재하는 임상시험인가? [ ](A-1.8, B-3 참조)
    - 무작위배정시험인가? [ ](A-1.8, B-5 참조)
    - 무작위배정시험이라면,
      - 각 군에 배당되는 수가 유사하여야 하는가? [ ](블록사용이 필요할 수 있다. B-5.6 참조)
      - 예후에 강력한 영향을 미치는 특정 요인이 존재하는가? [ ](층화 무작위배정이 필요할 수 있다. B-5.7 참조)
      - 예후에 영향을 미치는 특정 요인이 있다면,
        - 적은 수의 피험자수가 제시되고 있는가? [ ](B-5.8 참조)
      - 환자군집이 처리군에 무작위배정되는가? [ ](B-5.9 참조)
      - 교차시험설계인가? [ ](B-5.10b 참조)
      - 평가자 눈가림이 적용되는가? [ ](A-1.8, B-4 참조)
      - 연구 참여자 눈가림이 적용되는가? [ ](A-1.8, B-4 참조)
- 전향적 연구인가? [ ](A-2 참조)
- 전향적 연구라면,
  - 추적관찰 기간이 명기되었는가? [ ](A-2 참조)

### 연구 참여자(A-3 참조)

- 모집단에 대하여 기술하였는가? [ ]

- 연구대상이 연구에 적합한 이유를 설명하였는가? [ ]
- 연구 참여자가 선택되는 방법을 기술하였는가? [ ]
- 선정/제외기준을 명기하였는가? [ ]
- 피험자 수가 거절이나 중도탈락 고려하여 제시되었는가? [ ]

#### 변수의 유형(A-4 참조)

- 결과변수와 설명변수 모두에서 변수의 유형과 측정 척도가 제시되었는가? [ ](A-4.1, A-4.2 참조)
- 데이터의 수집방법에 대하여 기술하였는가? [ ](A-4.3 참조)
- 설문지나 표준화되지 않은 측정법을 사용하는 경우 신뢰도와 타당도에 대한 정보를 제시하였는가? [ ](A-4.4, A-4.4a, A-4.4b, A-4.4c 참조)

#### 피험자 수

- 피험자 수 산출 과정을 제시하였는가? [ ] (D-1 참조)
- 피험자 수 산출에 사용되는 결과 변수를 정의하였는가? [ ](D-5 참조)
- 임상적으로 의미 있는 효과 범위를 정의하였는가? [ ] (D-4.5 참조)
- 표본 수 산출 과정에서의 검정력과 유의수준을 기술하였는가? [ ](D-4.3, D-4.4 참조)
- 예상되는 응답률과 기타 피험자 감소 요인들이 고려되었는가? [ ] (D-6 참조)
- 제시된 피험자 수가 연구목적에 부합하는가? [ ] (D-7 참조)
- 피험자 수가 제시된 분석방법과 일치하는가? [ ] (D-7 참조)
- 피험자 수 산출 공식에 대한 기술이 적합한가? [ ] (D-8의 예제 참조)

#### 통계적 분석

- 적절한 용어를 사용하여 통계분석 방법을 기술하였는가? [ ](E-1.1, E-1.2 참조)
- 제시된 분석 방법이 연구에서 얻어지는 데이터의 유형에 적합한가? [ ](E-2, E-2.1, E-11 참조)
- 제시된 방법에 필요한 가정들이 만족되는가? [ ](E-4, E-4.1 참조)
- 제시된 방법이 데이터 셋의 구조(예를 들어 계층, 집단, 짝짓기, 짝지어진 데이터 등)를 고려하였는가? [ ](E-3, E-6, E-6.2, E-10 참조)
- 중요한 혼란변수들이 기술되고 그러한 변수들을 보정할 방법이 제시되었는가? [ ](E-5 참조)
- 다중 검정을 수행할 것인가? [ ](E-7.1, E-7.2, E-7.3, E-7.4, E-7.4a, E-7.4b, E-7.4c, E-7.4d, E-7.4e, E-7.4f 참조)
- 평균으로의 회귀 등의 측정오차로 인한 비뚤림이 고려되었는가? [ ](E-8 참조)
- 신뢰구간 산출에 관한 사항이 자세히 제시되었는가? [ ](E-12 참조)
- 의도된 치료법에 따른(ITT, intention to treat) 분석을 수행할 것인지의 여부와 합당한 이유를 제시하였는가? [ ](E-9 참조)

## 부록 2 용어 정리

한글	영문	위치
1-1로 짝지은 환자-대조군 연구	1-1matched case-control study	E-3
가짜치료	dummy treatment	B-3
개입	Intervention	B-3
개체	subject	E-3
검사-재검사 신뢰도	test-retest reliability	A-4.4b
검정력	power	B-5.9 B-5.10e D-3 D-4.3
결과변수	outcome variables	B-6
경향성에 대한 카이제곱 검정	chi-square test for trend	E-2
계층적 자료	hierachical data	E-6
계통할당	systematic allocation	B-5.4
과거 대조군	historical control	B-5.4
관찰적	observational	A-1.1 B-3
교란	confounding	A-1.5d A-1.6
교란변수	confounder	A-1.6
교란요인	confounding factor	A-1.6
교차설계	crossover design	B-5.10b
교차설계시험	cross over trial	A-1.7
교차시험	crossover trial	B-5.10c
교호작용	interaction	A-1.6a B-5.10e
구간척도	interval scale	A-4.1
군집	cluster	B-5.9
군집	clustering	E-6.1
그룹 내 비교	within group comparisons	B-5.10c
난수	pseudorandom numbers	B-5.5
난수표	tables of random numbers	B-5.5
눈가림	blind	A-1.8
눈가림	blindness	B-3 B-4
눈가림법	blinding	B-2
다기관 연구	multi-centre trials	B-5.5
다변량 분석	multivariate analysis	E-1.1
다수준 모형	multilevel modeling	E-6.1
다수준 자료	multilevel data	E-6
다중검정	multiple testing	B-5.10d
다중검정	multiple testing	B-6 B-7.1
다중요인분석	multifactorial analysis	B-6
단면적 연구	cross-sectional study	A-1.1 A-1.5
단변량 분석	univariate analysis	E-1.1
단일눈가림	single-blind	A-1.8 B-4.1
단일요인 분석	unifactorial analysis	E-1.1
대상 내 표준편차	within-subject standard deviation	A-4.4c
대응쌍	pair	E-3

한글	영문	위치
대조군	control	E-3
대조군	control group	A-1.4 B-3
데이터 모니터링	data monitoring	B-7
데이터 모니터링 위원회	data monitoring committee	B-7.1
독립표본 t 검정	two sample t test	E-2
독립표본 t 검정	unpaired t test	E-3 E-4
독립표본의 비교	unpaired comparison	E-3
동의	consent	B-8.1
로그변환	logarithmic transformation	E-4.1
명목척도	nominal scale	A-4.1
무작위배정되지 않은 개입 연구	non-randomized	D-3
	intervention study	
무작위배정 후 동의를 취득하는	randomized consent design	B-8.6
경우		
무작위 할당	random allocation	B-3 B-5.1 B-5.3
무작위배정	randomization	B-5.1 B-5.2 B-5.3
무작위배정 비교임상시험	randomized controlled	A-1.8
	clinical trial	
무작위배정표	randomization list	B-5.5
민감도	sensitivity	A-1.5b
민감도 분석	sensitivity analyses	B-9
민감도와 특이도 연구	study of sensitivity and	A-1.1
	specificity	
반복성	repeatability	A-4.4b
변동	variation	E-1.1
병동	hospital ward	C-1.1
병원	clinic	B-5.5
병행그룹	parallel group	B-5.10a B-5.10d
복합가설	composite hypothesis	E-7.4d
복합귀무가설	composite null hypothesis	E-7.3 E-7.4a
부가변수	additional variables	B-6
부작용	side effect	B-4.2
분율	proportion	D-3 D-8.1 D-8.2
분율	proportion	E-12.1
블록 무작위배정	randomization in blocks,	B-5.6
	block randomization	
비뚤림	bias	A-4.3 B-4.1 B-5.3 B-5.5
비활성약	inert drug	B-4.1
생태학적 연구	ecological study	A-1.1 A-1.5d
선정기준	inclusion criteria	B-1
설명변수	explanatory variable	E-1.1
세계 의학 협회	World Medical Association	B-3
순위척도	ordinal scale	A-4.1
순차적 설계	sequential design	B-5.10d
시험	trial	A-1.7 B-3
시험설계	trial design	B-2 B-5.10

한글	영문	위치
신뢰구간	confidence intervals	E-7.4 E-12
신뢰도	reliability	A-1.5e A-4.4b
신뢰성	confidence	D-2
실험	experiment	A-1.7 B-3
실험적	experimental	A-1.1 B-5.2
안전성	safety	B-3
양성예측도	positive predictive value (PPV)	A-1.5c
연구대상수	sample size	B-5.9 B-5.10d B-5.10e
연구정보를 충분히 제공한 동의	informed consent	B-8
연구책임자	trial organizer	B-8.6
연속형 자료	continuous data	A-4.2
예측변수	predictor measurement	E-7.4b
외삽	extrapolation	C-5
요인설계	factorial designs	B-5.10e
위약	placebo	B-3 B-4.2
위약효과	placebo effect	B-4.2
위양성	false positive	A-1.5c
위음성	false negative	A-1.5c
위험요인	risk factor	A-1.4
유병률 연구	prevalence study	A-1.1 A-1.5a
유의성 검정	significance test	E-5,6 E-7.1 E-7.3 E-7.4 E-11,12
유해사례	adverse events	B-7.1
유효성	efficacy	B-3
유효성	validity	E-4 E-12
윤리위원회	ethics committee	B-1
윤리적, 과학적 검토	ethical and scientific review	B-3
음성적 시험	negative trial	D-7
응급상황	emergencies	B-8.2
의도된 대로 분석법	intention to treat analysis	B-8.9
이분형 변수	binary variable	E-2
이산형 자료	discrete data	A-4.2
이중눈가림	double-blind	A-1.8 B-4.1
이중위약	double dummy	B-4.2
이차결과변수	secondary outcomes	B-6
이차종료점	secondary endpoints	B-6
일차결과변수	primary outcome	B-6
일차종료점	primary endpoint	B-6
일차진료기관	general practice	B-3 B-5.2
임계값	critical values	B-7.2
임계점	cut-off point	D-4.3
임상시험	clinical trial	A-1.7 B-3 B-5.1 B-5.2 B-5.3 B-5.6
임상시험윤리위원회	research ethics committees	B-8.1

한글	영문	위치
임상적으로 유의한 효과 크기	effect size of clinical importance	D-3 D-4.5 D-9
임상지침서	clinical guidelines	B-5.9
임의표본	random sample	B-5.3
잔차분산	residual variance	E-7.4d
장거리 배정	long range allocation	B-5.5
전향적	prospective	A-2
점검기록	audit trail	B-5.5
정규성 도표	normal plot	E-4
정량적 연구	quantitative study	D-1
정량적 자료	quantitative data	A-4.2
정밀도	precision	E-6.1
정밀성	precision	D-2
정성적 연구	qualitative research	D-1
정성적 자료	qualitative data	A-4.2
정확 방법	exact method	D-8.1
제곱근 변환	square root transformation	E-4.1
제외기준	exclusion criteria	B-1
주결과 변수	main outcome variable	B-6
주요인 효과	main factor effect	B-5.10e
집단 무작위배정	cluster randomization	B-5.2
집단 무작위배정설계	cluster randomized designs	B-8.5
짝지은 표본 t 검정	paired t test	E-3
짝지은 표본의 비교	paired comparison	E-3
짝지은 설계	matched pairs design	B-5.10c
참음성	true negative	E-7.2
최소유의차검정	least significant difference analysis	E-7.4d
최소화법	minimization	B-3 B-5.7 B-5.8
추적관찰	follow-up	A-2
축차설계 임상시험	sequential trial	B-8.10
층화 무작위배정	randomization in strata,	B-5.7
층화무작위배정	stratified randomization	E-5
치료법	treatment	B-2 B-3
치료의 잔류효과	carry-over of treatment effect	B-5.10b
카이제곱 검정	chi-square test	E-2
코호트 연구	cohort study	A-1.1 A-1.3
타당도	validity	A-1.5e A-4.4a
탐색적 연구	exploratory study	A-1.9
통제된 시험	controlled trial	B-3
특이도	specificity	A-1.5b
파일럿연구	pilot study	A-1.9
평가 비뚤림	assessment bias	C-2
평가자간 신뢰도	Inter-rater reliability	A-4.4c



한글	영문	위 치
평가자간 일치도	inter-rater agreement	A-4.4c
표본조사	sample survey	A-1.5
표본추출	sampling	D-2 D-9
표준편차	standard deviation	E-4
피험자 정보지	information sheets	B-8.1
피험자 정보지	patient information sheet	B-5.5
하위집단	subgroup	E-7.1
헬싱키 선언	declaration of Helsinki	B-3 B-8.1
환자군	case	E-3
환자 내 설계	within-patient design	B-5.10c
환자-대조군 연구	case-control study	A-1.1 A-1.4 C-1.1 C-1.2 C-2
활성약	active drug	B-4.1
황금기준	gold standard	A-1.5b
회상 비뚤림	recall bias	C-3
휴약기간	wash out	B-5.10b
Fisher의 정확검정	Fisher's exact test	E-2
Wilcoxon 분호 순위 검정	Wilcoxon signed rank test	E-3

## 참고문헌

1. 신영수, 안윤옥 편저 (1997) 의학연구방법론. 서울대학교출판부.
2. 안윤옥, 유근영, 박병주 (1999) 실용의학통계론. 서울대학교출판부.
3. 안윤옥, 유근영, 박병주, 김동현, 배종면, 강대희, 신명희, 이무송 (2005) 역학의 원리와 응용. 서울대학교출판부.
4. Altman DG (1982) Misuse of Statistics is unethical. In Statistics in Practice (Eds SM Gore and DG Altman). British Medical Association, London.
5. Altman DG. (1991) Practical Statistics for Medical Research. Chapman and Hall, London.
6. Altman DG, Bland JM. (1997) Units of analysis. British Medical Journal 314 1874.
7. Altman DG. & Bland JM. (1998) Generalisation and extrapolation. British Medical Journal 317 409-410.
8. Altman DG, Machin D, Bryant T, Gardner MJ. (2000) Statistics with Confidence, 2nd. ed., British Medical Journal, London.
9. Altman DG, Matthews JNS. (1996) Interaction 1: Heterogeneity of effects. British Medical Journal 313 486.
10. Armitage P, Berry G, Matthews JNS. (2002) Statistical Methods in Medical Research, 4th ed. Blackwell, Oxford.
11. Bland JM. (2000b) Sample size in guidelines trials. Family Practice 17 S17-S20.
12. Bland JM. (2002a) WMA should not retreat on use of placebos. British Medical Journal 324, 240.
13. Bland JM. (2002b) Fifth revision of Declaration of Helsinki: Clause 29 forbids trials from using placebos when effective treatment exists. British Medical Journal 324 975.

14. Bland JM and Altman DG. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*; i 307-10.
15. Bland JM and Altman DG. (1994) One and two sided tests of significance. *British Medical Journal* 309 248.
16. Bland JM and Altman DG. (1994a) Regression towards the mean. *British Medical Journal* 308 1499.
17. Bland JM and Altman DG. (1994b) Some examples of regression towards the mean. *British Medical Journal* 309 780.
18. Bland JM & Altman DG. (1994c) Matching. *British Medical Journal* 309 1128.
19. Bland JM, Altman DG. (1995) Multiple significance tests: the Bonferroni method. *British Medical Journal* 310 170.
20. Bland JM, Altman DG. (1996) Measurement error. *British Medical Journal* 313 744.
21. Bland JM, Altman DG. (1996a) Measurement error and correlation coefficients. *British Medical Journal* 313 41-2.
22. Bland JM, Altman DG. (1996b) Measurement error proportional to the mean. *British Medical Journal* 313 106.
23. Bland JM and Altman DG. (2002) Validating scales and indexes. *British Medical Journal* 324 606-607.
24. Bland JM, Kerry SM. (1997) Trials randomized in clusters. *British Medical Journal* 315 600.
25. Bland JM, Kerry SM. (1998) Weighted comparison of means. *British Medical Journal* 316 129.
26. Bland M. (2000) *An Introduction to Medical Statistics*, 3rd. ed. Oxford University Press, Oxford.

27. Bland M. (2000a1) An Introduction to Medical Statistics, 3rd. ed. Oxford University Press, section 9.10.
28. Bland M. (2000a2) An Introduction to Medical Statistics, 3rd. ed. Oxford University Press, section 10.7.
29. Breslow NE and Day NE. (1980) Statistical Methods in Cancer Research: Volume 1 – The analysis of case-control studies. IARC Scientific Publications No. 32, Lyon.
30. Breslow NE and Day NE. (1987) Statistical Methods in Cancer Research: Volume II – The design and analysis of cohort studies. IARC Scientific Publications No. 82, Lyon.
31. Burr ML (1993) Epidemiology of Asthma: in Burr ML (ed): Epidemiology of Clinical Allergy. Monogr Allergy Vol 31, Basel, Karger. p 80–102.
32. Collett D. (1994) Modelling Survival data in Medical Research. Chapman & Hall, London.
33. CLINSTAT software
34. Conover WJ. (1980) Practical Nonparametric Statistics, 2nd ed. John Wiley & Sons, New York.
35. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials.
36. Day SJ. & Altman DG. (2000) Blinding in clinical studies and other studies. British Medical Journal 321 504.
37. Edwards P., Roberts I., Clarke M., DiGiseppi C, Pratap S., Wentz R., Kwan I. (2002) Increasing response rates to postal questionnaires: systematic review. British Medical Journal 324 1183–1185.
38. Elashoff JD. (2000) nQuery Advisor Version 4.0 User's Guide. Los Angeles, CA.
39. Ferriman A. (2001) World Medical Association clarifies rules on placebo controlled trials. British Medical Journal 323 825.

40. Geleijnse JM, Witteman JCM, Bak AAA, den Breijen JH, Grobbee D E. (1994) Reduction in blood pressure with a low sodium, high potassium, high magnesium salt in older subjects with mild to moderate hypertension. *British Medical Journal* 309 436-40.
41. Gibbons, A. (2002) Performing and publishing a randomized controlled trial. *British Medical Journal* 324, S131.
42. Goldstein, H. (1995) *Multilevel Statistical Models*, 2nd ed. Arnold, London.
43. Guthrie E, Kapur N, Mackway-Jones K, Chew-Graham C, Moorey J, Mendel E, Marino-Francis F, Sanderson S, Turpin C, Boddy G, Tomenson B. (2001) randomized controlled trial of brief psychological intervention after deliberate self poisoning. *British Medical Journal* 323, 135-138.
44. Hayes RJ. (1988) Methods for assessing whether change depends on initial value. *Statistics in Medicine* 7 915-27.
45. Henshaw DL, Eatough JP, Richardson RB. (1990) Radon as a causative factor in induction of myeloid leukaemia and other cancers. *Lancet* 335, 1008-12.
46. Huskisson, E.C. (1974) Simple analgesics for arthritis. *British Medical Journal* 4 196-200.
47. Kerry SM, Bland JM. (1998) Analysis of a trial randomized in clusters. *British Medical Journal* 316 54.
48. Kerry SM, Bland JM. (1998b) Sample size in cluster randomization. *British Medical Journal* 316 549.
49. Kerry SM, Bland JM. (1998c) Trials which randomize practices I: how should they be analysed? *Family Practice* 15 80-83
50. Kerry SM, Bland JM. (1998d) Trials which randomize practices II: sample size. *Family Practice* 15 84-87
51. Kerry SM, Bland JM. (1998e) The intra-cluster correlation coefficient in cluster

- randomization. *British Medical Journal* 316 1455.
52. Lemeshow S, Hosmer DW, Klar J & Lwanga SK. (1996) Adequacy of sample size in health studies. John Wiley & Sons, Chichester.
53. Lilienfeld AM, Lilienfeld DE. (1980) Foundations of Epidemiology, 2nd ed.. Oxford University Press, Oxford.
54. Machin D, Campbell MJ, Fayers P, Pinol A. (1998) Statistical Tables for the Design of Clinical Studies, Second Edition Blackwell, Oxford.
55. Matthews JNS, Altman DG, Campbell MJ, and Royston P. (1990) Analysis of serial measurements in medical research. *British Medical Journal* 300 230-35. (Available online, but note that the link to the paper is the last of the bullet points.)
56. Matthews JNS, Altman DG. (1996a) Interaction 2: compare effect sizes not P values. *British Medical Journal* 313 808.
57. Matthews JNS, Altman DG. (1996b) Interaction 3: How to examine heterogeneity. *British Medical Journal* 313 862.
58. Nott MR, Peacock JL (1990) Relief of injection pain in adults – EMLA cream for 5 minutes before venepuncture. *Anaesthesia* 45 772-774.
59. Oldham PD. (1962) A note on the analysis of repeated measurements of the same subjects. *J Chron Dis* 15 969.
60. Pocock SJ. (1983) *Clinical Trials: A Practical Approach*. John Wiley and Sons, Chichester.
61. Rothman KJ, Michels KB, Baum M. (2000) For and against. Declaration of Helsinki should be strengthened. *British Medical Journal* 321 442-445.
62. Thomas M, McKinley RK, Freeman E, Foy C. (2001) Prevalence of dysfunctional breathing in patients treated for asthma in primary care: cross sectional survey. *British Medical Journal* 322, 1098-1100.
63. Tollman SM, Bastian H, Doll R, Hirsch LJ, Guess HA. (2001) What are the

effects of the fifth revision of the Declaration of Helsinki? British Medical Journal 323 1417-1423.

64. Vollmer WM. (1988) Comparing change in longitudinal studies: adjusting for initial value. J Clin Epidemiol 14 651-657.
65. Whitehead, J. (1997) The Design and Analysis of Sequential Clinical Trials, revised 2nd. ed. Chichester, Wiley.
66. Wetherill GB (1981) Intermediate Statistical Methods. Chapman & Hall, London.
67. Zelen, M. (1979) A new design for clinical trials. New Eng J Med 300 1242-5.
68. Zelen, M. (1990) randomized consent designs for clinical trials: an update. Statistics in Medicine 9 645-6.

## 연 구 진

### 연구책임자: 한서경

서울대학교병원 조교수

서울대학교 의과대학/서울대학교병원 의학연구협력센터 총괄팀장

### 공동연구자: 박병주

서울대학교 의과대학 예방의학교실 교수

서울대학교 의과대학/서울대학교병원 의학연구협력센터장

### 참여연구원

#### 최남경

서울대학교 의과대학 예방의학교실 박사과정

#### 김화정

서울대학교 의과대학 예방의학교실 박사과정

#### 이승미

서울대학교 의과대학/서울대학교병원 의학연구협력센터 역학연구원

#### 홍현숙

서울대학교병원 임상시험센터 통계연구원

#### 최혜미

서울대학교 의과대학/서울대학교병원 의학연구협력센터 통계연구원

#### 박지영

서울대학교 의과대학/서울대학교병원 의학연구협력센터 통계연구원

#### 진선영

서울대학교 의과대학/서울대학교병원 의학연구협력센터 통계연구원