

발 간 등 록 번 호

11-1470550-000104-01

의료기기 임상시험 관련 통계기법 가이드라인

2010. 8.

머 리 말

의료기기의 임상시험 계획 수립 및 평가 시 고려해야할 변수, 통계적 분석방법의 선택 등은 결과에 미치는 영향이 크기 때문에 올바른 자료 분석과 선택 방법 및 판단기준 설정은 매우 중요합니다. 특히 의료기기를 대상으로 하는 임상시험은 의약품과 비교하였을 때 대조군 설정의 어려움, 눈가림 유지 불가능 등의 많은 제약점이 따르기 때문에 의약품과 차별화된 임상 시험 예수 및 통계적 결과 분석 기법에 대한 가이드라인 제시가 필요합니다.

이에 식품의약품안전평가원에서는 2009년에 수행한 용역연구사업인 ‘의료기기 임상시험 관련 통계기법에 관한 연구’의 결과물을 토대로 의료기기의 임상 시험과 관련된 통계기법을 제시하고자 의료기기 임상시험 관련 통계기법 가이드라인을 발간하게 되었습니다.

본 가이드라인은 ‘의료기기 기술문서 등 심사에 관한 규정’에 대한 개정 (2010.06.11)내용을 반영하여 최근 인·허가 시 이용되는 임상시험 예수 및 통계적 결과분석기법을 자세히 설명하고 있어, 임상시험과 관련된 다양한 의료기기의 제품개발에 도움이 될 것으로 판단됩니다.

식품의약품안전평가원 의료제품연구부는 앞으로도 다양한 의료기기에 대한 평가 가이드라인을 지속적으로 발간해 나갈 예정이며, 이러한 가이드라인들이 제품개발 및 허가·심사 등 관련 업무 수행에 많은 도움이 될 수 있기를 기대합니다.

감사합니다.

2010년 8월

식품의약품안전평가원장

목 차

1. 의료기기 임상시험 고려사항	1
2. 주요 용어 설명	3
3. 통계적 디자인의 배치형태	9
4. 다기관 임상시험	14
5. 비교검증에 따른 분류	16
6. 비무작위 배정 임상시험 시 편향(bias) 보정	18
7. 연구대상자 수의 결정	19
8. 자주 사용하는 통계분석 방법 정리	27
9. 의료기기 임상시험 수행 중 고려사항	30
10. 의료기기 임상시험 자료의 통계분석 시 고려사항	32
11. 의료기기 임상시험에서 사용되는 통계분석법	37

1. 의료기기 임상시험 고려사항

☞ 의료기기의 임상시험 설계 시, 임상시험의 최종목표인 안전성과 유효성을 타당하게 증명해야 하며, 다음과 같은 사항에 대한 논의가 충분히 진행되어야 하고 구체적으로 기술되어야 한다.

1.1 임상시험계획서에 포함되어야 할 사항 중 통계적 방법을 고려해야 할 항목

- 의료기기법 시행규칙 제12조 제2항 참조

1. 임상시험의 명칭
2. 임상시험실시기관의 명칭 및 소재지
3. 임상시험의 책임자·담당자 및 공동연구자의 성명 및 직명
4. 임상시험용 의료기기를 관리하는 관리자의 성명 및 직명
5. 임상시험의뢰자의 성명 및 주소
6. 임상시험의 목적 및 배경
7. 임상시험용 의료기기의 사용목적 (대상질환 또는 적응증을 포함한다.)
8. 피험자의 선정기준·제외기준·인원 및 근거
9. 임상시험 기간
10. 임상시험 방법 (사용량·사용방법·사용기간·병용요법 등을 포함한다.)
11. 관찰항목·임상검사항목 및 관찰검사방법
12. 예측되는 부작용 및 사용 시 주의 사항
13. 중지·탈락 기준
14. 성능의 평가기준, 평가방법 및 해석방법 (통계분석방법에 의한다.)
15. 부작용을 포함한 안전성의 평가기준·평가방법 및 보고방법
16. 피험자동의서 서식
17. 피해자 보상에 대한 규약
18. 임상시험 후 피험자의 진료에 관한 사항
19. 피험자의 안전보호에 관한 대책
20. 그 밖의 임상시험을 안전하고 과학적으로 실시하기 위하여 필요한 사항

1.2 의료기기 임상시험 계획 수립 시 고려사항

1. 안전성 및 유효성을 밝히기 위해 구체화된 연구목표
2. Pilot or feasibility 연구 필요여부 또는 그 결과
3. 변수의 정의 및 선택
4. 연구 대상이 되는 모집단(Inclusion/Exclusion Criteria)
5. 타당한 대조군 설정 및 정의
6. 무작위 배정 방법
7. 가설에 따른 임상시험 디자인 종류
8. 눈가림의 방법 선택
9. 임상시험 실시 기관 및 연구자
10. 과학적으로 타당한 피험자 수 산출 근거 및 통계적 검정력
11. 성능평가 기준이 되는 일차/이차 효과 변수
12. 안전성 평가 대상 변수

2. 주요 용어 설명

☞ 임상시험계획 설계는 의료기기의 성능이 명확하고 객관적으로 측정될 수 있도록 기술되어야 한다. 이러한 목적을 위해서는 다음과 같은 용어가 먼저 명확하게 정의되어야 한다.

2.1 모집단

모집단이란 관찰의 대상이 되는 집단 전체로 연구 대상의 모집단은 실험이 시작되기 이전에 포함/불포함 기준을 명확하게 제시하고 전문 의료인의 판단하에 신중하게 결정되어야 한다. 또한 결정된 연구 모집단은 미래의 의료기기 사용자 집단과 비슷한 성격을 가져야 한다.

비슷한 집단에서 얻어진 시험 결과는 대상간의 변이가 작아지기 때문에 상대적으로 작은 표본 수를 가지고 임상시험이 진행될 수 있지만, 임상시험결과를 일반화하기 어렵다는 단점을 가지게 된다. 따라서 일반화를 시켜도 무리가 없을 정도로 연구 대상자의 특징이 고르게 분포된 집단을 선택하도록 해야 한다.

2.2 자료의 대표성

연구 대상 모집단이 미래에 의료기기를 사용할 모집단에 속한 환자들의 특성을 잘 나타낸다면 임상시험 결과는 유효하게 사용될 수 있다. 이런 특징을 가지는 표본을 대표성을 가진 표본이라고도 한다.

2.3 예비시험

임상시험 시 연구에 따른 경험이나 연구와 관련된 자료가 부족할 때에는 본 실험이 행해지기 전에 소규모의 예비실험 즉, 제한적인 사전 실험을

행할 수 있다. 예비실험의 효과 혹은 연구대상자에 미치는 육체적, 정신적 영향이 본 실험이 행해지기 전에 소멸되지 않는 경우에는 예비실험에 참여한 연구대상자는 본 실험에 참여해서는 안된다. 이런 예비실험에서는 의료기기가 가질 수 있는 문제점 혹은 부작용을 찾아낼 수 있을 뿐 아니라 실험 과정의 적합성, 신뢰할 수 있는 반응변수 측정 가능성 및 반응변수의 상한 한계 혹은 하한 한계 등을 추측할 수 있다. 예비실험에서는 제한적 가설검정이 가능하고 의료기기와 결과변수들 사이에 의미 있는 연관성을 찾기 위한 탐색적 자료 분석이 가능하다.

2.4 일차변수 및 이차변수

임상시험에서 밝히려고 하는 가장 중요한 목표(의료기기의 효과)를 정확히 나타내 줄 수 있는 변수를 일차변수로 정한다. 여러 측정 변수 중, 의학적 중요도 및 객관적 측정 가능성 등에 근거하여 구체적으로 변수를 선택한다. 기존의 임상시험이나 연구에서 이미 사용되어진 것으로써 연구자들에게 친숙한 변수를 사용할 것을 추천한다. 또한 일차변수는 의료기기 임상 시험의 연구대상자를 대상으로 측정이 가능한 신뢰할 수 있는 변수여야 하며 일차변수를 기반으로 표본수가 계산되어야 한다.

한 개의 변수로 일차목표를 측정하는 것이 불가능한 경우, 여러 측정변수들을 연계하여 복합 변수를 만들어 낼 수 있다. 이러한 연계 방법은 임상시험계획 단계에서 구체적으로 서술되어야 하며, 의학적 중요도에 비추어 변수가 해석 가능해야 한다. 복합 변수가 일차변수로 선택될 경우 복합변수를 구성하는 개별적 변수들의 재분석 결과도 추가로 제시하는 것이 좋다.

이차변수는 일차 목표 외에 추가적으로 보고 싶은 효과 측정을 위한 변수로, 이때 유효성이 입증되지 않더라도 일차 유효성 목표가 성공하면 임상시험 결과는 성공으로 본다. 다만, 이차 유효성 분석 결과를 통계적·의학적 지식

을 바탕으로 제시해야 한다.

2.5 공변량

공변량(covariates)은 임상시험결과 측정변수에 영향을 미칠 수 있는 연구대상자들의 특징을 설명하는 변수이다. 이러한 연구대상자들의 특징에 의해 발생할 수 있는 편향은 무작위 배정에 의해 줄일 수 있지만, 더 나은 보정을 위해 통계적 기법들을 사용하는 것을 권장한다. 특히 무작위 배정방법이 비슷한 공변량 값을 가지는 연구대상자들을 시험군에 고르게 배분하지 못한 경우 통계적 기법을 이용한 보정이 필요할 수 있다.

2.6 대조군

처리군과 대조군에 배정된 연구대상자들의 차이를 비교 분석한 결과로 의료기기의 안전성 및 유효성을 알 수 있다. 윤리적인 문제로 대조군 설정이 어려운 아주 드문 경우를 제외하고는, 대조군을 설정하는 것이 의료기기 임상시험의 경우 타당하며 일반적인 방법이다.

대조군은 다음과 같이 크게 4가지로 나눌 수 있다.

가. 동시(concurrent) 대조군: 같은 임상시험 연구책임자에 의해 같은 조건하에서 대립처리 시험이 행해지는 대조군으로 세 가지 종류의 대립처리 방법들은 다음과 같다.

- 1) 처리대조: 처리군에 상응하지만 다른 종류의 처리를 받은 경우를 말한다.
- 2) 무처리
- 3) 위약 또는 sham 기기

나. 수동적인 동시 대조군: 동시 대조군과 비슷하지만, 대조군의 시험이 같은 임상시험 연구책임자에 의해서 행해지지 않은 대조군을 의미한다.

다. 자기(self) 대조군 혹은 교차 대조군: 각 연구대상자가 시험군/대조군에 해당되는 치료 또는 처리를 받고 일정한 시간을 아무런 처리 없이 보낸 후에 대조군/시험군의 처리를 다시 받는 경우를 자기 대조군 혹은 교차대조군이라고 한다. 자기 대조군은 한 명의 연구대상자가 대조군, 처리군으로 동시에 사용되어서 연구대상자간의 변이를 설명할 수 있다는 것이 장점이지만, 대조군으로써 재측정 되기 전에 먼저 가해졌던 처리의 효과가 완전히 소멸될 때 까지 기다려야 하는 단점이 있다.

라. 시차 대조군: 같은 조건을 만족하지만, 시차 혹은 장소의 차이를 두고 행해진 시험 결과들을 모아서 만든 대조군을 말한다.

동시 대조군과 자기 대조군이 비교시험을 하기에 매우 용이한 방법들이다. 수동적 동시 대조군인 경우는 다른 시험자의 연구대상자 선택기준과 변수 측정 방법 등의 시험 여건들이 일치한다는 가정을 할 수 있을 때 유용하게 사용될 수 있다. 반면, 네 가지 대조군들 중에서 유용한 정보를 얻기에 가장 적합하지 않은 대조군은 시차 대조군으로 의료행위와 식생활의 시간적, 공간적 이질성 등과 같은 여러 가지 점검해야 할 사항들이 있지만, 이러한 차이를 정확히 측정하는 것은 현실적으로 어렵다.

대조군이 포함된 경우에 무작위 배정이나 눈가림법(blinding)은 편향을 줄이는 방법으로 유용하게 사용할 수 있다. 무작위 배정은 동시(concurrent) 대조군에서 흔히 사용되며 이러한 무작위 배정방법은 반드시 계획서에 명시되어야 한다. 그리고 신체에 이식되는 의료기기 임상시

험의 경우에는 연구자의 눈가림은 윤리적인 이유로 대부분 거의 불가능하다. 하지만 측정자의 눈가림은 가능하다.

2.7 처리할당방법

처리나 위약 등을 연구대상자들에게 할당할 때, 선택편향(selection bias)을 최소화 하도록 노력해야 한다. 시험결과와 연관성이 높은 예후인자(prognostic factor)를 가진 연구대상자들이 대조군에 많이 속하고 그렇지 않은 연구대상자들이 처리군에 많이 속하게 되는 경우에는 편향이 발생할 수 있다. 이러한 선택편향을 줄이기 위하여 가장 흔히 사용하는 방법이 무작위 배정(randomization) 방법 이다. 이 방법을 사용하면 연구대상자가 각 군에 할당될 확률은 동일하며 처치를 제외한 어떤 다른 변수에도 영향을 받지 않게 된다. 특히, 비교대상이 되는 그룹의 수가 적고 연구대상자의 수가 클 수록 균형있게 배분되며 연구자에 의한 의식적 또는 무의식적 편향도 제거 할 수 있다. 예를 들어, 무작위 배정방법이 없다면 의료전문가는 상대적으로 상태가 좋지 않은 연구대상자들에게 좀 더 강한 치료요법(처리)을 사용하려할 수도 있다. 반면, 연구대상자의 수가 적고 비교대상이 되는 그룹의 수가 많은 경우, 단순한 무작위 배정방법은 예후인자를 가진 연구대상자들을 균형있게 배분하지 못할 수도 있기 때문에 적절한 무작위 배정방법을 선택해야 한다.

2.8 눈가림(blinding)

눈가림(blinding)은 연구대상자, 시험진행자, 시험결과 측정자에 의해 발생하는 편향을 배제, 혹은 줄이기 위하여 사용 되는 방법이다. 연구대상자에게 처리에 관한 정보를 제공하지 않는 눈가림은 반응변수가 주관적으로 측정될 때 특히 중요하다. 예를 들어, 시술 후 연구대상자의 고통완화 정도를 측정하는 경우 연구대상자의 눈가림은 측정의 객관성을 높일 수 있다. 시험진행자에게 처리에 관한 정보를 제한하는 눈가림은 처리의 할당, 자료 분석의

객관성을 증대시킨다.

편향을 최소화 하기 위해 다음과 같은 네 가지 눈가림법 들이 사용될 수 있다.

가. 한쪽 눈가림법: 시험진행자나 연구대상자 중 한쪽에만 눈가림법이 행해지는 방법이다. 일반적으로 많이 행해지는 한쪽 눈가림법은 시험진행자가 시험의 처리에 관한 정보를 제공받지만 연구대상자는 처리에 관한 정보를 제공받지 못하는 방법이다. 이 경우, 연구대상자는 어떠한 처리가 자신에게 행해지는지 모르기 때문에 위약효과(placebo effect)를 배제할 수 있다. 하지만 시험진행자가 처리에 관한 정보를 가지고 있기 때문에 연구대상자들에 따라 다른 처리방법을 사용할 수 있고, 처리에 관한 정보를 연구대상자에게 유출시킬 가능성이 있다는 한계점이 있다.

나. 양쪽 눈가림법: 시험 진행자와 연구대상자 모두에게 연구대상자가 받는 처리를 모르게 하며 제3의 측정자가 연구대상자에게 어떠한 처리가 행해지는지를 결정하고 기록하는 방법이다. 이러한 경우, 위약효과나 처리에 관한 정보 유출을 배제할 수 있어 객관적인 결과를 가져올 수 있다.

다. 제 3자 눈가림법: 시험진행자와 연구대상자는 어떠한 처리를 받는지 알지만 측정자에게는 모르게 하는 방법이다. 의료기기 임상시험의 경우 처치방법을 연구대상자나 시험진행자에게 눈가림 하는 것이 현실적으로 매우 어려운 경우가 많으며, 이 경우 사용할 수 있다.

라. 삼중 눈가림법: 시험의 구성방법에 따라 의미는 변할 수 있지만 일반적으로 연구대상자, 시험진행자 뿐 아니라 측정자까지 어떠한 처리가 사용되고 있는지를 모르게 하는 방법이다. 때로는 시험 결과에 대한 객관성을 높이기 위해 측정자에 대한 눈가림이 시험진행자에 대한 눈가림보다 더 중요할 수 있다.

3. 통계적 디자인의 배치형태

☞ 의료기기 임상시험에서 통계적 시험설계는 비교시험에서 대조군과 시험군의 배치 형태에 따라 1. 평행설계(parallel design), 2. 교차설계(crossover design), 3. 요인설계(factorial design)로 구분할 수 있다.

3.1 평행설계 (Parallel Design)

평행설계는 의료기기 임상시험뿐만 아니라 대부분의 비교 임상시험에서 가장 일반적인 형태를 가진 설계방법으로, 연구대상자는 무작위 배정에 의해서 서로 다른 처리군(2집단 또는 3집단 이상)으로 배정이 된다. 연구의 종료시까지 처음 배정된 군을 유지하며 진행된다.

이 설계는 다른 설계에 비해 단순한 배치 체계를 가지고 있는 것이 장점이지만, 연구진행 중의 약정위배, 중도탈락, 결측자료, 공변량의 보정, 반복측정에 따른 교호작용 등 발생할 수 문제점들에 대해 대비책을 준비하여야 하므로 분석은 복잡해 질 수 있다.

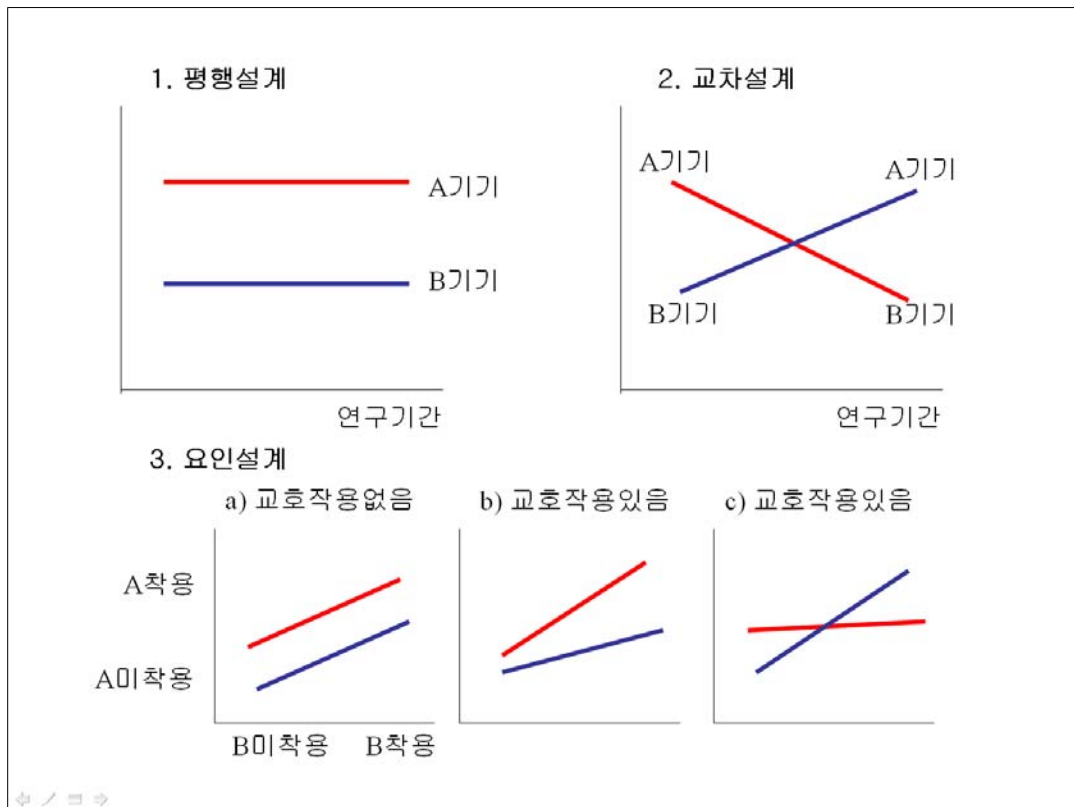


그림 1 통계적 임상시험 설계의 예시

3.2 교차설계 (Cross-Over Design)

교차설계는 평행설계와 다르게, 한 연구대상자에게 시험용 의료기기, 대조용 의료기기 모두를 적용하여 측정하는 방법이다. 예컨대 대조군과 시험군이 나누어진 경우(2×2 교차설계), 피험자는 대조 의료기기 우선군 및 시험의료기기 우선군으로 나뉘는 무작위 배정을 받게 된다. 즉, 대조 의료기기 우선군에 배정된 경우에는 대조 의료기기를 먼저 사용(착용)하고 이전 의료기기의 효과가 사라질 정도의 일정시간 후에 시험의료기기를 착용하게 되며, 시험의료기기 우선군에 배정될 경우에는 시험의료기기를 먼저 사용(착용)하고, 일정 시간 후 대조의료기기를 착용하게 되는 경우가 된다(그림 1). 결과적으로 한명의 연구대상자에서 대조군과 처리군을 직접비교할 수 있다는 장점이 있다. 또한, 한명의 연구대상자에게서 두 번의 측정치를 얻게 되므로 부수적으로 연구대상자의 수가 평행설계보다 줄어드는 장점도 가지게 된다.

교차설계에서는 연속적으로 두 기기의 시험을 수행하기 때문에 첫 번째 순서에서 수행한 의료기기의 효과인 잔류효과(residual or carryover effect)가 발생하지 않도록 충분한 시간(wash out period)를 두고 시험을 진행해야 한다. 교차설계는 1) 적응증의 특성이 급성질환으로 자주 재발하고, 2) 치료 의료기기가 일시적 증상 완화/치료하는 질환에 적합하며, 3) 연구대상자의 탈락이 상대적으로 적게 일어날 것이라고 예측되는 질환에 관련된 의료기기 임상시험에 적용하는 것이 알맞다.

3.3 요인설계 (Factorial Design)

요인설계는 두 개 이상의 처리군의 조합에 대해서 동시에 효과를 검정할 수 있는 유용한 방법으로 특정 처리군의 조합을 통해 상호작용(interaction)을 확인할 수 있는 통계적 디자인이다. 가장 단순한 예인 2×2 요인설계에서 한 요인으로 A 의료기기를 적용한 경우와 그렇지 않은 경우, 또 다른 한 요인으로 B 의료기기를 적용한 경우와 그렇지 않은 경우를 고려하면 이 두요인의 조합은 각각 (A, B 의료기기 적용하지 않은 경우), (A 의료기기 적용, B 의료기기 비적용), (A 의료기기 비적용, B 의료기기 적용), 그리고 (A, B 의료기기 적용)의 4가지 경우가 나타나게 된다. 이와 같은 상황에서 두 의료기기가 모두 효과가 있음을 확인한 검정은 다음과 같이 비교를 통해서 가능할 뿐만 아니라, A, B 의료기기를 동시에 사용하였을 경우에 상승작용이 있는지를 확인할 수 있다.

- 1) (A, B 의료기기 비적용) vs (A 의료기기 적용, B 의료기기 비적용)
- 2) (A, B 의료기기 비적용) vs (A 의료기기 비적용, B 의료기기 적용)

그림 1의 3(a)의 경우는 서로 상호작용이 없는 경우이며, (b)의 경우는 긍정적인 상호작용이 있는 것의 예이며, (c)의 경우는 각각의 두 요인의 효과가 반감하여 작용하는 상호작용이 있는 예가 된다.

요인설계는 두 의료기기 이상의 복합치료나 의료기기 및 의약품의 복합치료에 대한 임상시험연구 시 효과 분석에 용이한 통계적 디자인이다. 요인설계에서는 두 요인간의 상호작용을 동시에 고려할 수 있기 때문에 적은 연구 대상자의 수를 통해서 이를 확인할 수 있는 장점이 있다.

평행설계와 요인설계의 예시는 다음과 같다.

1) 평행설계

그룹A	그룹B	그룹C
무처리군	항생제투여군	개인용조합자극기군

2) 요인설계

항생제투여	개인용조합자극기처리여부	
	비처리	처리
항생제비투여	A 그룹	C 그룹
항생제투여	B 그룹	D 그룹

3.4 배치 형태에 따른 설계의 선택

지금까지 배치형태에 따라 크게 3가지 유형(평행설계, 교차설계, 요인설계)의 배치형태를 확인할 수 있었는데, 임상시험 설계 시, 배치 유형의 판단은 다음과 같다.

가. 평행설계 vs 교차설계

평행설계 또는 교차설계를 선택하는 판단기준으로 ① 해당 질병의 특성이 급성으로 자주 발생하는지, ② 연구 의료기기가 질병의 일시적 호전/완화에 기여하는지, ③ 연구대상자의 중도탈락이 많은지 않은지, 그리고 ④ 해당 연구 의료기기의 이상반응이 국소적인 경우인지에 따라 모든 경우를 만족할 경

우 교차설계를 고려할 수 있다. 만약 위의 4가지 기준중 하나라도 만족하지 못한 경우에는 평행설계를 고려할 수 있다.

나. 평행설계 vs 요인설계

평행설계 또는 요인설계를 선택하는 판단기준으로는 ①고려되는 연구 의료기기가 둘 이상인지, ②연구 의료기기를 동시에 적용하였을 경우 상호 작용이 더 중요한 관심사항인지 여부와 같은 사항이 있다.

4. 다기관 임상시험

☞ 임상시험에서 다기관 시험을 고려하는 이유는 다음과 같다. 첫째로 여러 기관에서 동시에 환자를 모집하기 때문에 임상시험에 소요되는 시간을 절약할 수 있다는 점과 두 번째로 다기관 임상시험에서 얻어진 결과는 일반화된 방법으로 인정을 받을 수 있다는 점 때문이다.

품목(변경)허가를 위한 안전성·유효성 심사로 활용하는 임상시험일 경우, 「의료기기기술문서등심사에관한규정(식품의약품안전청 고시 제2010-42호)」에 의거하여 하나의 적응증마다 2개소 이상의 임상시험실시기관(해당 의료기기의 특성과 임상시험방법 등을 종합적으로 고려하여, 임상시험예수가 통계적으로 타당하게 결정되어야 함)에서 실시되어야 하며, 이 경우 의료기기 임상시험 실시기관은 「의료기기임상시험실시기관지정에관한규정(식품의약품안전청 고시 제2010-15호)」에 따라 식품의약품안전청장으로부터 의료기기 임상시험 실시기관으로 지정받은 곳이어야 한다.

4.1 의료기기 다기관 임상시험 시 고려사항

의약품 임상시험에서와 다르게 의료기기 임상시험에서 다기관 임상시험이 주요하게 다뤄지는 이유는, 의료기기의 유효성은 의료기기 본연의 효과뿐만 아니라 임상상의 의료기기 시술 과정(특히 침습적 기구의 삽입)이 의료기기 성능에 영향을 끼치는 경우가 많기 때문이다. 일반적으로 연구자의 시술능력과 치료 효과가 서로 교란(confounded)되어 있어 의료기기 임상시험 실시기관 간의 차이에 대한 임상시험결과 역시 차이가 나기 마련이다.

따라서 다기관 임상시험을 고려할 때에는 분석을 통하여 기관 간 안전성 유효성 결과차이에 대한 분석이 분명히 이루어져야 한다. 만약 실시기관간 차이가 발생할 경우에는 그 사유에 대한 심층분석이 이루어져야 하며, 실시기관과 상호작용을 하는 요인에 대한 탐색이 요구된다.

실시기관간의 차이를 여부를 확인하는 통계적 방법으로 처리효과를 고정효과로 설정하고, 실시기관의 효과, 처리와 실시기관의 상호작용을 임의효과로 고려하는 일반화 혼합모형이 있다. 가장 기본적인 (선형)모형의 식은 다음과

같다.

$$y = \mu + \alpha_i + \gamma_j + \alpha\gamma_{ij} + \varepsilon$$

여기서 α_i 는 처리효과(고정효과), 즉 처리군별 차이이고 γ_j 는 실시기관효과(임의효과), $\alpha\gamma_{ij}$ 는 처리효과와 실시기관효과의 상호작용(임의효과)이다.

이 뿐만 아니라 subgroup 분석에서는 유효성 안전성 평가 결과뿐만 아니라 실시기관과 일반적 사항, 질병 예후 등에 관련 변수를 면밀히 분석하여 실시기관간 차이여부를 확인하고, 만약 이에 대한 차이가 있을 경우 이를 보정할 수 있는 통계적 모형들(층화분석, 보정을 위한 회귀 또는 선형 모형, 짝짓기, 성향점수 방법 등)을 고려하여 보정된 처리효과를 평가하는 것을 권고한다.

5. 비교검증에 따른 분류

☞ 비교검증의 기준에 따른 종류를 구분하여 보면 의료기기 임상시험은 우월성 비교, 비열등성 또는 동등성 검정 그리고 제한적으로 반응-효과 검정으로 구분할 수 있다.

5.1 우월성 비교

우월성 비교는 연구목표가 대조군(sham 또는 active 의료기기)에 비하여 처리군(연구 의료기기)의 효과가 더욱 뛰어남을 보이려는 목적이 있을 경우에 해당한다. 그러나 연구 참여자의 질병이 심각하거나 sham 대조군을 고려한 우월성 비교는 윤리적으로 문제가 발생할 수 있다. 이때는 활성대조군을 고려하여야 한다.

5.2 비열등성 또는 동등성 비교

대조군의 유효성과 유사하거나 열등하지 않음을 입증하려는 방법은 비열등성 또는 동등성 비교에 해당한다. 비열등성 검정은 통상적인 대조군(활성 대조군이여야 함)에 비하여 치료/진단 효과가 처리군에서 열등하지 않음을 보이는 것이 되고, 동등성 비교는 유효성의 평가 결과가 대조군과 유사한 경향(열등하지도 않고, 우월하지도 않음)을 보이는 경우가 된다.

일반적으로 동등성과 비열등성은 활성대조군과 처리군의 효과 차이에 대한 신뢰구간을 이용하여 쉽게 해석되는데, 그림 2는 동등성 판정의 기준을 제시하여 주고 있다. 만약 미리 정해진 임상적으로 무시할 수 있다고 판단될 차이를 δ (margin)라고 할 때, 동등성과 비열등성은 활성대조군과 처리군의 효과 차이에 대한 신뢰구간이 $\pm \delta$ 의 범위안에 포함되어 있는지에 대한 평가가 가능하다. 그림 2의 검정색 화살표는 바로 동등성이 입증된 예가 되고, 붉은 색의 경우 동등하다고 할 수 없는 상황이며, 파란색은 동등성을 판

단할 수 없는 상황이다. 비열등성은 활성대조군과 처리군의 효과 차이에 대한 신뢰구간의 단측 만을 확인하여 평가할 수 있다.

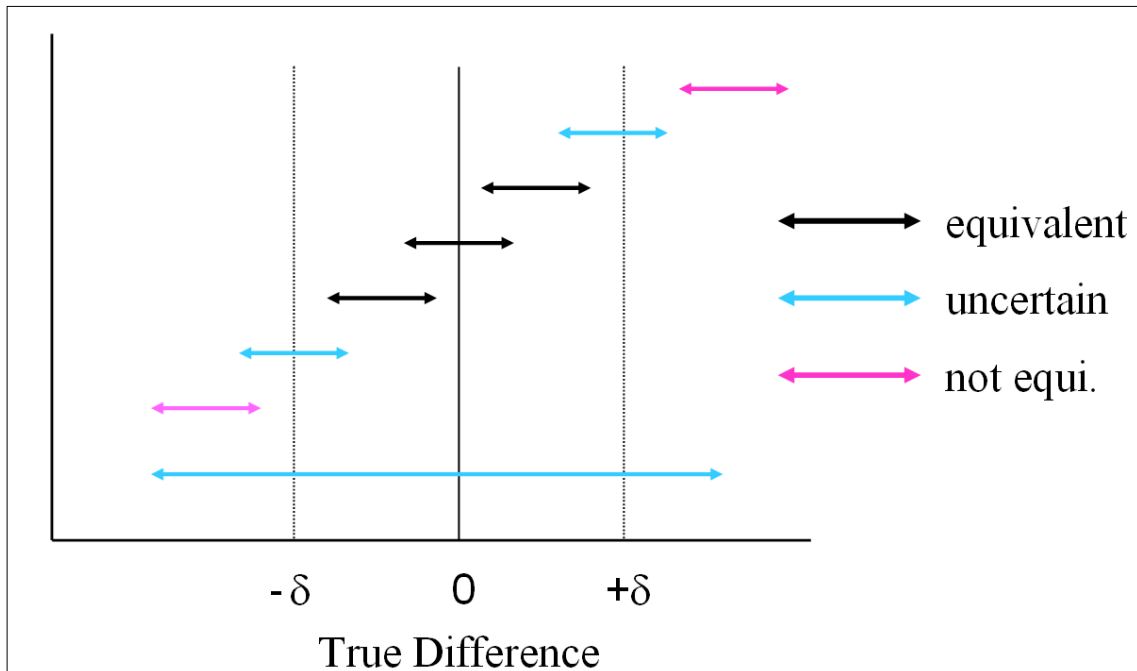


그림 2 동등성(비열등성)의 판단

동등성 및 비열등성 검정을 수행하기 위해서는 임상적으로 무시할 만한 차이, 즉 한계점(margin)에 대한 결정이 연구계획 단계에서 이루어져야 한다. 무시할 만한 차이는 통계적인 측면뿐만 아니라 임상적인 측면에서 동등성 및 비열등성 검정의 성패를 결정짓는 중요한 요인이 된다. 이를 결정하는데 일반적인 규칙은 정립된 것이 없지만 “적어도 sham군의 효과와 활성대조군의 효과 또는 예상되는 연구 의료기기의 효과의 차이의 절반이하여야 한다”는 일반론적인 규칙이 존재한다. 그러나 sham 효과가 작은 유효성 변수를 고려할 경우 치료효과와 sham 효과간의 차이가 크게 벌어져 있기 때문에 이 원칙을 그대로 적용하기에는 무리가 따르는 경우도 있다.

활성 대조군의 선정에 있어서도 매우 주의를 기울여야 하는데, 가장 대표적인 기준으로 현재 임상에서 가장 널리 사용되고 있고, 표준 치료라고 고려되는 의료기기 또는 진단 방법을 기준으로 삼아야 한다. 즉, 이러한 활성대조군의 연구가 잘 정립되어 있어 선행 우월성 검정의 자료들이 충분한 것을 대조군으로 설정해야 한다.

6. 비무작위 배정 임상시험 시 편향(bias) 보정

☞ 무작위 배정 비교 임상시험이 아닌 비무작위 배정 임상시험을 수행하였을 경우에는 편향의 탐색 및 보정이 필수적이다.

6.1 비열등성 또는 동등성 비교

교란변수(confounder)에 의한 편향 보정방법으로 설계단계에서는 짝짓기(matching) 방법, 분석단계에서는 층화(stratification) 또는 회귀분석(regression analysis)을 통한 교란변수 보정을 이용할 수 있다. 이런 방법을 통하여 교란변수 보정 전의 결과와 보정 후의 결과를 비교하여 차이가 있다면 그 원인을 밝히도록 노력해야 한다.

7. 연구대상자 수의 결정

☞ 임상연구에서 연구대상자의 수는 연구목표를 분석할 수 있을 정도의 충분한 수가 보장되어야 한다. 1차 주효과 변수를 기준으로 판단해야 하며, 임상시험대상 피험자 수의 결정 방법 및 근거가 기술되어야 한다. 그러나 최근 임상연구가 점차 복잡해짐에 따라, 다양한 연구목적에 평가하기 위하여 1차 주효과 변수 뿐만 아니라 2차 효과 변수와의 조합 등의 방법이 분석에 고려되고 있다.

7.1 연구 대상자 수 산출을 위한 필요정보

피험자 수를 결정하기 위해서는 필요한 사전 정보는 다음과 같다.

- 1) 연구가설
- 2) 유의수준
- 3) 통계적 검정력
- 4) 사용될 통계적 분석 방법(즉, 연구디자인과도 관련됨)
- 5) 선행연구 또는 문헌 리뷰를 통한 예상되는 효과 차이(및 표준편차)

1)-5)를 알아보기 위해서는 통계적 가설 검정 메커니즘을 이해해야 한다. 연구 목표를 정하면, 이를 구체적인 가설로 만들어 설정하여야 한다. 가설은 다음과 같이 크게 두 종류로 구분된다.

가. 대립가설 (H_1): 연구자가 임상시험을 통해 증명하려는 가설

나. 귀무가설 (H_0): 대립가설의 입증에 실패하여 채택할 수밖에 없는 가설

다음으로, 유의수준과 검정력에 대한 설정이 필요하다. 이 유의수준과 검정력은 표 1에 정리된 통계적 가설검정에서의 오류를 범할 확률로 표현 된다.

표 1 통계적 가설검정의 결정에서 나타나는 오류의 종류

Reality Decision \	H ₀ is True	H ₁ is True
H ₀ is True	Correct	Type II error ^{**}
H ₁ is True	Type I error [*]	Correct (Power)

* Type I error: 제 1종 오류 **Type II error: 제 2종 오류

표 1의 의사결정 과정에서 유의수준 (α)이란 제 1종 오류를 범할 최대허용 한계치이며, 귀무가설이 참인 경우 귀무가설을 기각하는 오류를 저지를 확률을 의미한다. 검정력 ($1 - \beta$)는 제 2종 오류, 즉 대립가설이 참인 경우 귀무가설을 기각하는 올바른 결정을 내릴 확률을 의미한다. 일반적으로 80%이상은 유지하며 좀 더 신뢰할 만한 결과를 얻기 위해서는 90%의 검정력을 유지하는 표본수를 구하도록 한다. 연구대상자 수의 결정 공식은, 연구디자인, 주효과 변수의 종류, 분석방법에 따라 다양한 형태로 구성되어 있기 때문에 이에 대한 전문적인 지식이 요구된다. 연구대상자의 수 산출에 있어 중요한 요건 중 하나는 선행연구 또는 문헌 리뷰를 통해 얻어진 ‘예상되는 치료 효과의 차이’ 및 표준편차에 대한 정보의 획득인데, 선행연구를 통한 근거 자료 및 과학적으로 타당한 설명이 반드시 이루어져야 한다.

7.2 표본수 계산 공식

표본수가 너무 작다면 의료기기의 효과를 파악해낼 수 있는 검정력이 상대적으로 낮아지는 문제점이 있다. 반대로 너무 많으면 지나치게 많은 시간과 자원을 소모하고, 비윤리적이며, 환자들이 더 좋은 처치를 받을 수 있는 기회를 박탈하는 연구가 될 것이다. 따라서 최적의 표본수를 계산한 뒤 연구를 수행하는 것은 매우 중요하다.

<용어정리>

μ_t : 시험기기에서 주효과변수의 평균

μ_c : 대조기기에서 주효과변수의 평균

p_t : 시험기기에서의 성공률

p_c : 대조기기에서의 성공률

d : 효과의 크기, 즉 $d = \mu_t - \mu_c$

σ : 주효과변수의 표준편차

ε : 동등성/비열등성 인정한계

α : 유의수준

$1 - \beta$: 검정력, 즉 β 는 제 2종의 오류를 범할 확률

n_t : 시험기기군의 연구대상자 수

n_c : 대조기기군의 연구대상자 수

7.2.1 독립 일표본의 평균 비교(우월성)

가설: $H_0: d=0$ vs $H_A: d \neq 0$

임상적인 최소 유효크기를 $(\mu_A - \mu_0) = d$, 제 2종 오류를 β 라 하면, 필요한 표본수는 다음과 같다.

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{d^2}$$

단, 여기서 모분산 σ^2 은 알려져 있다라고 가정한다.

예

새로운 안압약 T의 효과를 검정하기 위하여 대조약 C를 이용한 교차 시험을 수행하였다. 즉, 각 연구대상자에게 무작위로 한쪽 눈에는 T 약을 투여하고 다른 쪽 눈에는 C약을 투여하여 안압의 강하의 차이가 있는지를 검정하려고 한다. 이 시험에서 얻어지는 자료는 짝지은 자료이며 양쪽눈의 안압의 강하의 차이의 평균이 0이 아닌지를 검정하는 단일표본 t-검정을 이용할 수 있다. 이 때 예상되는 군간 차이는 10mmHg이며 차이의 표준편차 추정치는 20mmHg으로 알려져 있다. 유의수준은 5%로 검정력은 최소 90%를 만족하는 최소 표본수는 위의 공식을 이용하면

$$\begin{aligned} n &= \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{d^2} \\ &= \frac{(1.96 + 1.282)^2 20^2}{10^2} \\ &\simeq 43 \end{aligned}$$

43명의 연구대상자가 계산된다. 10%의 탈락율을 고려하면 총 48명의 표본이 필요하다.

7.2.2 독립 이표본의 평균 비교(비열등성)

주효과변수의 평균과 미리 정해진 값과의 차이가 인정한계 ε 보다 작지 않거나(비열등성)/ 크다는 것을 검정하는 경우에 사용하는 방법. 즉, 가설이 다음과 같은 경우 (여기서 μ_c 는 미리 주어진 값이다)

$$H_0: \mu_t - \mu_c \leq \varepsilon \quad \text{vs} \quad H_A: \mu_t - \mu_c > \varepsilon$$

표본수는 다음과 같다.

$$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{(|d| - \varepsilon)^2}$$

단, 모분산 σ^2 가 근거자료에 의해 타당하게 제시되어 있음

예 전립선 비대증 연구대상자에서 레이저기기A를 이용하여 비대해진 전립선을 기화시켜 배뇨기능을 개선시킨 효과가 기존 레이저기기B의 효과보다 비열등함을 보고자 하는 연구를 진행하려고 한다. 주효과 변수는 베이스라인으로부터 시술 12주 후 국제전립선증상점수(연속형변수임)의 평균변화정도이다. 현재 기기가 기존 레이저 기기의 미리 알려진 평균변화정도인 9.7보다 못하지 않다는 가설검정을 위한 표본수는 다음과 같다. 단, 변화량의 표준편차(σ)는 7.3으로 알려져 있고 비열등성 한계(ε)는 -2.5로 가정하기로 한다.

위의 값들을 고려하여 유의수준(α) 5%, 검정력($1-\beta$)를 80%로 설정한 표본수는 다음과 같다.

$$H_0: \mu_t - \mu_c \geq \varepsilon \quad vs \quad H_A: \mu_t - \mu_c < \varepsilon$$

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{[(\mu_t - \mu_c) - \delta]^2} = \frac{(1.642 + 0.842)^2 \times 7.3^2}{[-2.5]^2} = 52.61$$

즉, 53명이 필요하다.

7.2.3 독립 이표본의 평균 비교(우월성)

$$\text{가설: } H_0: \mu_t - \mu_c = 0 \quad vs \quad H_A: \mu_t - \mu_c \neq 0$$

가. 군당 표본수가 같은 경우

$$n = \frac{2\sigma^2(z_{\alpha/2} + z_\beta)^2}{d^2}$$

나. 군당 표본수가 다른 경우

$$n_t = \frac{(\sigma_t^2 + \sigma_c^2/r)(z_{\alpha/2} + z_{\beta})^2}{d^2}, \quad n_c = rn_t$$

여기서 $r = n_c/n_t$ 이다.

7.2.4 독립 이표본의 평균 비교(비열등성)

$$\text{가설: } H_0: \mu_t - \mu_c \leq \varepsilon \quad \text{vs } H_A: \mu_t - \mu_c > \varepsilon$$

가. 군당 표본수가 같은 경우

$$n = \frac{2\sigma^2(z_{\alpha} + z_{\beta})^2}{(|d| - \varepsilon)^2}$$

나. 군당 표본수가 다른 경우

$$n_t = \frac{(\sigma_t^2 + \sigma_c^2/r)(z_{\alpha} + z_{\beta})^2}{(|d| - \varepsilon)^2}, \quad n_c = rn_t$$

여기서 $r = n_c/n_t$ 이다.

7.2.5 두 비율(성공율) 비교(우월성)

$$\text{가설: } H_0: p_t - p_c = 0 \quad \text{vs } H_A: p_t - p_c \neq 0$$

가. 군당 표본수가 같은 경우

$$n = \frac{(z_{\alpha/2} \sqrt{2 \bar{p} \bar{q}} + z_{\beta} \sqrt{p_1 q_1 + p_2 q_2})^2}{d^2},$$

여기서 $d = p_t - p_c$, $\bar{p} = (p_1 + p_2)/2$, $\bar{q} = 1 - \bar{p}$

나. 군당 표본수가 다른 경우

$$n_t = \frac{(z_{\alpha/2} \sqrt{(r+1) \bar{p} \bar{q}} + z_{\beta} \sqrt{r p_t q_t + p_c q_c})^2}{r d^2}, \quad n_c = r n_t$$

여기서 $\bar{p} = \frac{p_t + r p_c}{r+1}$, $r = n_c / n_t$ 이다.

7.2.6 두 비율(성공율) 비교(비열등성)

가설: $H_0: p_t - p_c \leq \varepsilon$ vs $H_A: p_t - p_c > \varepsilon$

가. 군당 표본수가 같은 경우

$$n = \frac{(z_{\alpha} \sqrt{2 \bar{p} \bar{q}} + z_{\beta} \sqrt{p_t q_t + p_c q_c})^2}{(|d| - \varepsilon)^2}$$

나. 군당 표본수가 다른 경우

$$n_t = \frac{(z_{\alpha} \sqrt{(r+1) \bar{p} \bar{q}} + z_{\beta} \sqrt{r p_t q_t + p_c q_c})^2}{r(|d| - \varepsilon)^2}, \quad \bar{p} = \frac{p_t + r p_c}{r+1}, \quad n_t = r n_c$$

7.2.7 생존분석 - log-rank test를 이용한 1:1 표본수인 경우

위험율은 다음과 같이 정의된다.

$$h = \frac{\ln(\pi_t)}{\ln(\pi_c)} ,$$

이 때 π_i 는 그룹 i 의 t 시점에서의 생존율이다.

표본수는 다음과 같이 계산된다.

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 (h+1)^2}{(2 - p_t - p_c)(h-1)}$$

8. 자주 사용하는 통계분석 방법 정리

☞ 해당 임상시험 설계방법 및 반응변수 종류에 따라 알맞은 통계분석 방법을 사용하도록 한다.

8.1 자주 사용하는 통계분석 방법 정리

가. 독립적인 수치형 반응변수

상황	통계분석법
단일집단 검정	1) 모수적 방법: 단일표본 t-검정 (one sample t-test) 2) 비모수적 방법: 윌콕슨 순위부호 검정 (Wilcoxon signed rank test)
독립적인 두 집단 비교	1) 모수적 방법: 독립 이표본 t-검정 (two sample t-test) 2) 비모수적 방법: 윌콕슨 순위합 검정 (Wilcoxon rank sum test) 또는 맨-위트니 U 검정 (Mann-Whitney U-test)
독립적인 세 집단 이상 비교	1) 모수적 방법: 분산분석 (analysis of variance: ANOVA) 및 다중비교 (multiple comparison) 2) 비모수적 방법: 크루스칼-윌리스 검정 (Kruskal-Wallis test)
연속형 독립변수와의 관계	1) 독립변수가 하나인 경우: 단순회귀분석 (simple linear regression) 2) 독립변수가 여러 개인 경우: 다중회귀분석 (multiple linear regression)
연속형과 범주형 독립변수와의 관계	1) 공분산분석 (analysis of covariance) 2) 일반선형모형 (general linear models)

나. 상관된 수치형 반응변수

상황	통계분석법
짝진 자료	1) 모수적 방법: 짝진 t-검정 (paired t-test)
	2) 비모수적 방법: 차이에 관한 윌콕슨 순위부호 검정 (Wilcoxon signed rank test on differences)
3회 이상 반복측정된 자료와 독립변수와의 관계	1) 다변량분산분석 (multivariate analysis of variance)
	2) 반복측정 분산분석 (repeated measures analysis of variance)
	3) 선형혼합모형 (linear mixed models)

다. 독립적인 범주형 반응변수

상황	통계분석법
하나의 범주형 독립변수와의 관계	1) 셀의 기대빈도가 모두 5 이상인 경우에는 카이제곱 검정 (chi-square test)
	2) 그렇지 않은 경우는 피셔의 정확검정 (Fisher's exact test)
	3) 반응변수가 이분형(binary)인 경우는 단순 로지스틱 회귀 분석 (simple linear regression)도 가능
두 개 이상의 독립변수 (범주형/연속형 모두 가능)와의 관계	1) 반응변수가 이분형인 경우는 다중 로지스틱 회귀분석
	2) 반응변수가 세 개 범주 이상인 경우는 multinomial or ordinal logistic regression

라. 상관된 범주형 반응변수

상황	통계분석법
짝진 범주형 자료분석	맥니마 검정 (McNemar's test)
3회 이상 반복측정된 범주형 자료와 독립변수와의 관계	1) 일반화 추정방정식(generalized estimating equation: GEE)을 이용한 주변모형 (marginal model)
	2) 일반화 선형 혼합모형 (generalized linear mixed model)

마. 기타 분석방법

상황	통계분석법
두 연속형 변수간의 상관	1) 피어슨 상관계수 (Pearson correlation coefficient) 2) 스피어만 상관계수 (Spearman's correlation coefficient)
생존자료 분석	1) 카플란-마이어 (Kaplan-Meier) 2) 로그-순위 검정 (log-rank test) 3) 콕스의 비례위험모형 (Cox's proportional hazard model) 등.
일치도 분석	1) 민감도 (sensitivity), 특이도 (specificity) 2) 카파값 (Kappa measure), 3) 일치상관계수 (concordant correlation coefficient) 4) ROC 분석 5) 블랜드-알트만 분석 (Bland-Altman analysis)

9. 의료기기 임상시험 수행 중 고려사항

☞ 임상시험 수행 시 시험이 완료되기 전 중간에 아래와 같은 사항을 고려해야 한다.

9.1 모니터링

임상시험이 계획에 따라 잘 수행되고 있는지를 파악하고, 계획 시 고려하지 못한 문제점 등을 파악하는 과정으로 중요한 역할을 하게 된다. 임상시험에 있어서 다음과 같은 문제가 발생할 수 있는 상황들을 모니터링 해야 한다.

- 가. 환자의 모집은 프로토콜에서 제안된 것과 같이 진행되고 있는지의 여부
- 나. 임상시험과정에서의 환자의 안전성
- 다. 통계적으로 적합한 처리의 할당
- 라. 변수측정의 빈도 및 정확성
- 마. 임상자료의 기록 및 보안

환자의 안전성문제를 다루는 경우에는 데이터 모니터링 위원회가 구성되어 정기적인 모니터링이 필요하다. 데이터 모니터링 위원회에는 경험을 가진 연구자를 최소한 한 명이상 포함시킬 것을 추천한다. 임상시험 모니터링의 주요방법은 임상시험과정 체크와 신뢰할 수 있는 자료가 수집되고 있는지를 확인하는 단순 모니터링과 일부 얻어진 자료에서 치료 효과를 중간에 체크하는 중간분석으로 나누어진다.

9.2 중간분석

중간분석(interim analysis)이란 환자가 모집되는 과정 혹은 임상시험 진행 도

중에 임상시험 센터의 수행능력, 수집된 자료의 정확성, 처리의 유효성 등을 확인하여 임상시험이 예상된 대로 잘 진행되는지 점검하는 것이다.

임상시험과정에서의 환자 안전성이나 임상시험 이후 심각한 건강상의 문제를 일으킬 수 있는 경우, 중간분석이 필요하다. 이때, 연구계획에 미리 중간 분석의 횟수, 분석방법, 중간분석으로 인한 추가 대상자 수 등이 기술되어야 한다.

10. 의료기기 임상시험 자료의 통계분석 시 고려사항

이 절에서는 임상시험에서 모아진 자료의 분석을 수행하는 경우에 고려해야 할 통계적 사항에 관하여 기술한다. 즉, 분석에 사용하는 분석군, 결측자료 및 이상치의 처리, 자료변환, 통계적 분석방법 등에 관하여 살펴본다.

10.1 분석군

분석에 포함될 연구대상자 집단은 임상시험 계획서에 미리 명확하게 정의해야 한다. 또한 분석처리 대상자의 기초정보(인구통계학적 정보, 질병상태 등)를 파악해야 한다. 임상시험에서 무작위 배정된 모든 연구대상자가 임상시험을 완벽하게 계획대로 순응한다고 볼 수 없기 때문에 사전에 문제점들이(시험계획서 위반, 중도탈락, 처치의 비순응 등) 분석 결과에 어떻게 영향을 주는지 명시해야 한다. 무엇보다 임상시험계획서에는 이러한 문제가 발생하는 빈도를 줄이고, 문제점을 고려하는 분석 방법이 제시되어야 한다.

이렇게 여러 가지 문제점들로 무작위 배정이 손상을 입은 경우에는 편향을 가장 최소화하는 방향으로 분석군을 정해야 한다. 현재 임상시험에서 가장 많이 고려되고 있는 분석군으로는 Intention to Treat (ITT)군과 Per Protocol (PP)군이 있다. ITT군은 간략히 말하면 연구대상자에게 어떤상황이 벌어졌더라도 애초 무작위 배정된 상태를 그대로 유지한 분석군이다. 이에 비해 PP군은 무작위 배정 후, 계획서대로 임상시험을 따르지 않은 연구대상자를 분석에서 제외한 군을 말한다. 현재 임상시험에서는 (특히, 우월성 검정을 위한 임상시험) ITT군을 주분석군으로 가장 많이 이용한다. 그 이유는 ITT군이 무작위 배정의 이점을 잘 살리고 있고 일반적으로 보수적인 결과를 주기 때문이다.

10.1.1 ITT(Intention to Treat) 분석군

배정된 대로 분석인 ITT 원칙은 무작위 배정된 모든 연구대상자를 처음 무작위 배정된 그대로 유지시켜 분석 한다는 원칙이다. 무작위 배정된 모든 연구대상자인 ITT군 이라는 용어는, 무작위 배정된 모든 연구대상자를 분석에 포함시키는 배정된 대로의 분석 원칙을 가장 근접하고 완전하게 적용할 수 있는 분석대상 집단을 뜻한다. 분석에서 최초의 무작위 배정상태를 그대로 유지하는 것은 표본의 편향발생을 최소화할 수 있다.

10.1.2 PP(Per Protocol) 분석군

PP군 분석은 임상시험계획서대로 시험을 완료한 연구대상자만 분석하는 방법이다. 즉, 의료기기 처치에 노출되고, 측정이 가능하며, 자격 요건에 맞고, 다른 중요한 계획서 위반을 하지 않는 대상자들을 의미한다. PP군 분석 접근은 처치의 최대의 잠재적 효과를 보려는 목적에 사용될 수 있으며, 일반적으로 널리 사용되는 ITT 분석과 비교분석 결과를 제시한다. PP군을 사용하면 처치의 효과가 과대평가될 수 있는 가능성이 크다. 즉, 처치에 따라 환자들의 순응도가 달라질 수 있는 경우 분석 결과에 편향이 있을 수 있다.

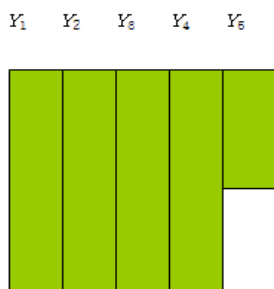
10.2 결측자료

임상시험에서 결측값(missing value)은 편향을 발생시키는 원인이 된다. 따라서 자료 수집 및 관리에 있어, 임상시험계획서에서 요구하고 있는 모든 사항을 충족시키기 위해 노력해야 한다. 자료의 결측을 최소화 하기 위해서 계획된 방문을 유도하는 등, 지속적인 연구대상자관리가 필요하다.

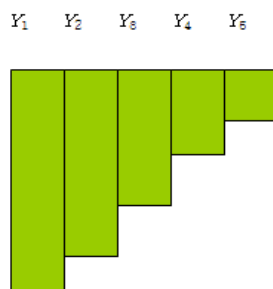
임상시험결과 분석 시 주 분석군을 ITT군으로 정의한 경우는 결측이 발생하게 되면 ITT의 원칙을 지킬수 없게 된다. 따라서 이러한 결측값을 포함한 자료를 분석하는 합리적인 방법을 사전에 미리 계획서에 제시해야 한다. 결측값 보정방법 및 결측값에 대한 가정에 따라 분석 결과는 달라진다. 따라서 상황에 따른 상이한 분석결과에 대해 검토하는 것 또한 중요하다.

결측의 패턴은 단변량패턴, 단조패턴, 일반패턴으로 구분된다. 단변량패턴은 측정된 여러 변수 중에서 단지 한 변수에서 결측이 발생하는 경우이다. 단변량 패턴의 대표적인 예는 회귀분석 자료에서 독립변수는 모두 측정되었으나 종속변수에서 결측이 발생한 경우이다(그림 3 (가)). 단조패턴은 임상시험에서 중도탈락이 일어난 경우에 가장 흔하게 일어난다. 연구대상자가 임상시험 수행 중에 중도 탈락되어 어느 시점 이후에는 결과변수의 관측값이 없는 경우가 해당된다(그림 3 (나)). 일반패턴은 일반적인 연구에서 가장 흔히 발생하는 결측패턴이다(그림 3 (다)).

(가) 단변량 패턴



(나) 단조패턴



(다) 일반패턴

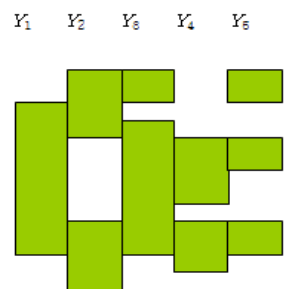


그림 3 결측자료 패턴

10.3 결측자료 분석법

10.3.1 완전히 관찰된 개체만을 이용한 분석법 (Complete Case Analysis)

이 방법은 모든 변수들이 관측된 개체들만 이용하여 분석하는 방법이다. 여러 변수 중에 단 하나의 변수에만 결측값이 있어도 그 개체는 분석에서 제외한다. 이 방법은 편향을 야기하지는 않지만 가지고 있는 자료의 누락이나 결측이 많은 경우는 매우 비효율적이다. 하지만 분석 방법이 매우 간편하고 일변량 통계들의 비교가 가능한게 장점이다.

10.3.2 우도 근거 방법 (Likelihood-based Analysis)

이 방법은 결측을 고려한 우도를 찾고 그 우도로부터 최대우도 추정치를 얻는 방법이다. 예를 들면, 선형혼합모형(linear mixed model 또는 mixed model for repeated measures)의 사용 분석이 우도 근거 방법이다. 선형혼합모형을 사용하면 가지고 있는 자료를 전부 이용하여 분석할 수 있으며, 이 때 임의결측인 경우 편향을 최소화하여 추정값을 얻을 수 있는 장점이 있다.

10.3.3 단일 대체법 (Single Imputation)

각 결측값을 하나의 값으로 대체하여 완전한 자료 행렬로 만들고 이 대체된 값들을 실제로 여기고 분석을 하는 방법이다. 이때 통계적 모형을 이용하여 대체값(imputed value)을 구한다. 몇 가지 자주 사용되는 단일 대체방법은 다음과 같다. 물론 이 외에도 여러 가지 대체모형에 의한 대체법을 고려할 수 있다.

가. 비조건부 평균 대체법 (Unconditional Mean Imputation)

: 완전히 관찰된 개체들을 이용하여 결측변수의 평균을 구한 후, 그 결측값을 그 평균으로 대체하는 방법이다.

나. 조건부 평균 대체법 (Conditional Mean Imputation 또는 Regression mean imputation)

: 결측값을 가진 변수를 종속변수로, 결측이 없는 다른 변수들을 독립변수로 하여 완전히 관찰된 개체들을 이용하여 회귀분석한 후, 결측이 없는 독립변수값들을 회귀식에 대입하여 결측값의 예측값을 구하고 대체하는 방법이다.

다. 마지막 관측 값 선행 대체법 (Last Observation carried Forward: LOCF)

: 경시적 자료(longitudinal data)에서 각 개체 내 결측값은 마지막으로 관측된 값으로 모두 대치한다. 예를 들어, 어느 한 임상시험이 1년간 시행

되며 시험참여자는 베이스라인(0개월), 3, 6, 9, 12개월 총 5번 방문하여 혈압을 측정한다고 가정하자. 이때, 한 시험참여자가 6개월까지만 혈압을 측정하고 그 이후 임상시험에서 중도탈락했다. 6개월에 측정된 혈압값은 150mmHg라고 하면 이 연구참여자의 9개월, 12개월 결측값은 모두 150mmHg로 대체된다. 이런 방법을 LOCF라고 한다. 이 방법은 미국 FDA의 오래된 지침으로 임상시험에서 가장 흔하게 사용되는 결측자료 대체방법이다. 이 방법이 임상시험에서 선호된 이유는 처치의 효과가 보수적으로 추정된다는 점인데 요즘 수행된 연구결과들은 LOCF가 보수성을 보장하지는 않는다고 밝히고 있고, 자료의 평균과 공분산 구조가 매우 왜곡되며 심지어 완전임의결측 가정에서도 편향이 발생할 수 있다고 주장하고 있다. 특히 의료기기 임상시험의 경우 LOCF는 보수성을 보장하지 않는 경우가 많기 때문에 분석방법 선택 시 신중해야 한다.

10.4 이상치의 처리

자료에 이상치(outlier)가 있는 경우 분석결과에 그릇된 영향을 줄 수 있다. 특정한 변수값을 이상치로 정의하기 위해서는 통계적·의학적 관점에서 자료에 대한 해석이 필요하며, 시험계획서에 이상치의 기준 및 처리에 관한 적절한 방법을 제시해야 한다. 이때, 이상치에 대한 처리 방법이 어느 의료기기를 선호하는 쪽으로 사전에 제시되어서는 안 된다. 사전에 계획서에 이상치에 대한 처리 방법을 규정하지 않을 경우에는 실제 자료값을 이용한 분석과 이상치를 제거 또는 감소시킨 최소한 하나 이상의 분석의 실시 및 그 결과가 비교 검토 되어야 하며 결과에 차이가 나는 경우 그 이유를 자세히 기술하여야 한다.

11. 의료기기 임상시험에서 사용되는 통계분석법

☞ 통계분석법은 보통 측정된 변수를 어떻게 해석할 것인가에 따라 여러 가지 다른 방법을 사용할 수 있다. 다시 말해, 연구자의 목적 및 변수의 종류에 따라 분석 방법이 달라진다.

11.1 연속형 자료 - 단일 집단의 평균 비교

변수가 연속형 자료이며, 그 집단의 평균과 임계값을 비교하는 방법이다. 단일 집단의 사람들로부터 얻어진 표본이 있고, 연구의 관심이 변수의 모평균이 어떤 특정한 값(임계값)과 같은 지 검정하는 방법이다. 이 경우 사용할 수 있는 분석방법은 일표본 t -검정(one sample t -test)과 비모수적 방법인 순위검정(sign test) 또는 윌콕슨 부호순위 검정(Wilcoxon's signed rank test)등 이 있다.

11.1.1 일표본(단일표본) t -검정

t -검정은 주 결과변수의 모평균 μ 가 미리 주어진 값 μ_1 과 차이가 있는지를 검정할 때 사용되는 방법이다. 이 검정은 자료가 독립이고 평균이 μ 이고 분산이 σ^2 인 정규분포를 따른다고 가정한다. t -검정의 귀무가설과 대립가설은 다음과 같다.

$$H_0: \mu = \mu_1 \text{ vs. } H_a: \mu \neq \mu_1$$

t -검정은 정규분포의 모표준편차 σ 를 알 수 없으므로 모분산 대신 표본분산 s^2 를 이용하여 다음과 같은 검정통계량을 계산한다.

$$t = \frac{(\bar{x} - \mu_1)}{(s/\sqrt{n})} \sim t_{n-1}$$

검정통계량 t 는 귀무가설 하에서 자유도가 $n-1$ 인 스튜던트 t 분포를 따르게 되므로, 이 분포를 이용하여 p -값을 구하고 유의수준과의 비교를 통해 검정을 실시한다. 표본수가 큰 경우 t 분포는 표준정규분포에 근사하므로 t 검정 대신 표준정규분포에 근거한 방법을 사용할 수도 있다.

11.1.2 부호검정(sign test)

분포의 평균을 사용하는 모수적인 방법과 달리, 데이터의 중앙값을 사용하기 때문에 정규성 같은 가정들이 만족되지 않을 때에도 사용할 수 있는 분석 방법으로, λ 보다 큰(혹은 작은) 표본의 개수를 검사하는(따라서 자료의 부호 " λ 보다 크면+ 작으면 -"만을 조사하는) 검정 방법이다.

n 개의 표본 중 λ 와 같은 값을 가지는 표본을 제외하고 남은 n' 개의 표본 중 λ 보다 큰 표본의 개수와 작은 표본의 개수를 비교하여 이 둘 중 작은 수를 r 이라 하고, $n' \leq 10$ 이면 r 자체가 검정 통계량이 되며, $n' > 10$ 인 경우에는 다음의 검정통계량 방법을 통해 부호검정을 하게 된다.

$$Z = \frac{\left| r - \frac{n'}{2} \right| - \frac{1}{2}}{\frac{\sqrt{n'}}{2}} \sim Normal(0,1)$$

이 방법은 매우 단순한 검정 방법이며, 좀 더 검정력이 높은 방법인 월콕슨의 부호순위 검정을 사용할 수도 있다. 이 방법의 경우 자료의 부호와 더불어 순위(rank)도 같이 고려하는 방법이다.

11.2 수치형 자료 - 서로 관련이 있는 집단

서로 관련이 있는 두 집단 또는 한 개체 내에서 측정된 두 수치형 자료는 서로 상관이 있다. 예를 들면, 무릎관절 치료에 사용되는 두 개의 의료기기

A와 B의 효과를 비교하기 위해 100명의 표본을 이용하는 임상시험이 있다고 하자. 이 때 시험 방법은 각 연구대상자에게 무작위로 A를 한 쪽 무릎에 B는 반대쪽 무릎에 적용하였다고 가정해 보자. 즉, 각 연구대상자에게서 A의 결과와 B의 결과를 모두 얻을 수 있다. 이 경우 연구대상자간은 서로 독립이지만 연구대상자 내에서 얻어진 A의 결과와 B의 결과는 서로 상관되어 있다.

11.2.1 짝진 t -검정 (paired t -test)

해당 모집단에서 각 차이값들은 특정 분산을 가지는 정규분포에 따르며, 이 표본 수는 정규성 검토를 할 수 있을 정도로 충분하다는 가정 아래에서 사용할 수 있다.

두 집단에서의 측정값들이 평균적으로 동일하다면 차이값의 평균은 0이 될 것이므로, 가설은 아래와 같이 세울 수 있으며

$$H_0: \mu_d = 0 \text{ vs } H_a: \mu_d \neq 0$$

통계량은 $t = \frac{\bar{d}-0}{s_d/\sqrt{n}} \sim t(n-1)$ 로, 일표본 t -검정과 같은 방식으로 검정을 할 수 있다.

11.2.2 윌콕슨의 부호-순위 검정(Wilcoxon's signed rank test)

부호 검정과 같은 조건 하에서 쓰일 수 있으며, 차이값에 대한 부호 뿐만 아니라 순위도 같이 사용하여 검정하기 때문에 부호 검정보다 검정력이 큰 방법이다. 중앙값을 기준으로 부호를 결정하여 자료를 크기 순으로 정리해서 번호를 매긴 다음, 부호가 양수인 순위들의 합(T_+)과 음수인 순위들의 합

(T_{-})이 같은지의 여부를 검정할 수 있다. 두 합 중 작은 값을 T 라고 하고, 중위수와 같은 표본을 제외한 n' 의 수에 따라, $n' \leq 25$ 이면 T 자체가 검정 통계량이 되며, $n' > 25$ 이면, 다음의 정규 근사를 통해 검정을 하게 된다.

$$z = \frac{\left| T - \frac{n'(n'+1)}{4} \right| - \frac{1}{2}}{\frac{n'(n'+1)(2n'+1)}{24}} \sim Normal(0,1)$$

11.3 수치형 자료- 서로 독립적인 두 집단

서로 관련이 없는 독립적인 두 집단의 수치형 혹은 순서형 변수를 비교하는 경우, 이 변수의 모평균 혹은 분포가 두 집단 간에 서로 차이가 있는지 파악하고자 할 때 사용된다. 예를 들어, 새로운 창상 피복제와 기존의 제품 중 한 쪽에 무작위 할당된 두 집단의 화상 환자들의 피부 회복 면적 등을 비교하고 싶은 경우 등에 쓰일 수 있다.

11.3.1 독립 이표본 t -검정 (Two independent sample t -test)

모집단에서 해당 변수는 집단별로 정규분포를 따르며, 두 집단의 분산은 서로 동일하고, 표본수 역시 정규성 가정 및 등분산 가정을 검토 할 수 있을 만큼 충분히 클 때 사용한다. 두 집단 간의 평균 차이를 비교하고자 하며, 두 집단의 모평균이 동일하다는 귀무가설이 참이라면, 이 차이는 0이 될 것이며가설은 다음과 같다.

$$H_0: \mu_1 = \mu_2 \text{ vs. } H_1: \mu_1 \neq \mu_2$$

따라서 검정통계량은 두 표본평균 간의 차이 및 귀무가설 하에서 정의되는 모

평균간의 차이에 근거해서 계산한다. 이 검정통계량은 t -분포에 따르게 되며 t 로 표기한다. 이때, 두 집단 간 결합 표준편차(pooled standard deviation)를 사용하게 된다.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{SE(\bar{x}_1 - \bar{x}_2)} = \frac{(\bar{x}_1 - \bar{x}_2)}{s_{pooled} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2) ,$$

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

또한 차이에 대한 신뢰구간의 범위에 따라 임상적인 유의성을 판단할 수도 있다. 따라서, 비록 t -검정 결과가 통계적으로는 유의하다 할지라도 임상적으로는 사실상 무의미한 차이가 될 수도 있다.

11.3.2 윌콕슨 순위합 검정(Wilcoxon's rank sum test)

보통 독립 이표본 t 검정의 가정이 만족되지 않을 때 많이 쓰이는 비모수적인 방법으로 맨-휘트니의 U 검정(Mann-Whitney's U-test)과 동등한 검정 방법이며 두 집단의 각각의 값들의 순위를 합한 것을 사용하는 검정 방법이다. 모든 관찰값들이 하나의 집단에서 얻어졌다는 가정 하에 두 집단의 자료를 섞어서 순위를 매긴다. 단, 같은 크기를 가지는 관찰값들은 평균 순위로 책정한다. 표본수가 작은 집단에 해당하는 관찰값들의 순위합(T)를 계산하여 검정에 이용하고, 만일 두 집단의 표본수가 모두 15 이하이면 T 가 검정통계량이 되며, 최소한 한 집단의 표본수가 15보다 클 경우 정규 근사를 사용해 검정통계량을 이용할 수 있다.

$$z = \frac{T - \mu_T}{\sigma_T} \sim N(0, 1) , \mu_T = \frac{n_s(n_s + n_L + 1)}{2} , \sigma_T = \sqrt{\frac{n_L \mu_T}{6}}$$

이때, n_s 는 작은 표본의 표본 수, n_L 은 큰 표본의 표본 수가 된다.

11.4 범주형 자료- 단일집단

n 명의 사람들로 구성된 단일 표본이 있고, 각 개체들이 연구 관심 특성을 보유하고 있는 경우(남자, 사망, 유병 등)와 그렇지 않은 경우로 나뉜다. 이와 같은 자료의 적절한 요약 통계량은 해당 특성을 보유하고 있는 사람들의 비율이다. 즉, 모집단의 실제비율이 어떤 특정값에 해당하는지 파악하는 것이다.

11.4.1. 단일 비율(π)에 관한 검정(test of single proportion)

표본 내 각 개인들은 해당 특성을 보유 하거나 보유하고 있지 않다는 가정하에 해당 특성을 가지는 사람들의 수는 이항분포에 의해 $n\pi$ 로 계산되며, 이때 샘플의 수 n 과 확률 π 의 곱이 5보다 크다면 정규분포로 근사시킬 수 있고, 평균의 추정치는 표준 편차 p 는 다음과 같다.

$$\sqrt{\frac{p(1-p)}{n}}$$

검정 가설은 다음과 같다.

$$H_0: \pi = \pi_1 \text{ vs. } H_1: \pi \neq \pi_1$$

아래의 검정통계량을 이용하여 위의 가설을 검정한다.

$$Z = \frac{(|p - \pi_1| - 1/2n)}{\sqrt{p(1-p)/n}}$$

π 에 대한 95% 신뢰 구간을 통해 중요성을 파악할 수도 있다. π 에 대한 95% 신뢰 구간은 다음과 같이 계산된다.

$$p \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$$

신뢰 구간이 넓게 되면 낮은 정밀도를 가진 추정치라는 의미가 되기도 한다.

11.5 범주형 자료- 두 비율 비교

서로 독립적인 집단이나 한 집단에서 특성에 관한 비율이 이들 두 집단에 차이가 있는지 없는지 여부를 조사하고자 할 때 쓰는 방법이다.

11.5.1 서로 독립인 두 집단 - 카이제곱 검정(Chi-square test, χ^2 -test)

가. 분할표 (contingency table) - 자료의 빈도 형태. 즉, 각 표본에서 해당 특성을 보유하고 있거나 보유하고 있지 않은 인원 수 등과 같은 내용을 각각의 칸에 빈도로 나타나게 된 아래와 같은 형태를 분할표라 한다.

표 2 2×2 분할표

특 성	집 단 1	집 단 2	합 계
있 음	a	b	$a+b$
없 음	c	d	$c+d$
합 계	$n_1=a+c$	$n_2=b+d$	$n=a+b+c+d$
특성을 가진 비율	$p_1=a/n_1$	$p_2=b/n_2$	$p=(a+b)/n$

각각의 빈도 값은 관측값이며, 이를 통해 귀무가설 하에서의 기대빈도 (expected frequency)를 계산한다. 즉, 집단1에서 특성을 가질 기댓값은 n_1p 이며, 집단 2에서 특성을 가질 기댓값은 n_2p 이다.

나. 가정 : 표의 각 행과 열은 상호 배타적(mutually exclusive)이다. 즉, 사람들은 표본에 한 번만 나타나고, 따라서 각 사람들은 단지 하나의 행 및 하나의 열에만 속하게 된다. 또한 최소한 하나의 칸에서 기대빈도가 5보다 작으면 카이제곱 검정을 사용하지 않고 피셔의 정확검정(Fisher's Exact test)을 사용하는 것이 좋다.

다. 이론 : 각 칸(cell)의 기댓값과 관찰값을 이용하여 검정통계량을 계산한다. 검정통계량은 다음과 같으며,

$$\chi^2 = \sum \frac{(O - E - 1/2)^2}{E}$$

여기서 O 와 E 는 각각 관찰값과 기댓값이 된다. 이 검정통계량은 귀무가설 하에서 자유도가 1인 카이제곱 분포를 따르게 되어 p -값을 구하여 검정을 하게 된다. 단일 비율과 마찬가지로 신뢰구간을 통해 결과를 내리기도 한다. 두 집단간 모비율 차이에 대한 95% 신뢰구간은 다음과 같다.

$$(p_1 - p_2) \pm 1.96 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

11.5.2 서로 관련이 있는 두 범주형 자료 비교 : 맥니마 검정(McNemar's test)

가. 배경 : 서로 다른 상황에서 동일한 사람들을 대상으로 같은 특성을 두 번 측정하는 경우 등과 같이 서로 관련이 있거나 또는 종속적이 두 집단에서 얻은 자료에서 각 사람들은 해당 특성을 두 상황 모두 가지고 있거나, 어느 한쪽만 가지고 있거나, 아니면 둘 다 가지고 있지 않은 경우 등으로 분류 된다(표 3).

표 3 짝진 이분형 자료의 분할표

특 성		상 황 1		합 계
		있 음	없 음	
상 황 2	있 음	w	x	$w+x$
	없 음	y	z	$y+z$
합 계		$w+y$	$x+z$	$n = w+x+y+z$

나. 다음의 검정통계량은

$$\chi^2 = \frac{(|x-y|-1)^2}{x+y}$$

자유도가 1인 카이제곱 분포를 따르며 분자의 1은 연속성 수정 (continuity correction)을 위한 값이다. 이 값을 분포표의 값과 비교하여 검정한다. 95% 신뢰구간은 다음과 같다.

$$\frac{x-y}{m} \pm \frac{1.96}{m} \sqrt{x+y - \frac{(x-y)^2}{m}}$$

11.6 상관 분석 (correlation analysis)

두 변수 X 와 Y 간의 선형 관련성 정도를 측정하기 위한 분석방법이다. 산점도(scatter plot)를 이용하면 이들 변수를 이차원 적으로 표시할 수 있어 두 변수의 관계를 눈으로 확인 할 수 있다.

11.6.1. 피어슨 상관계수(pearson's correlation coefficient)

관찰된 점들의 중간을 가로지르는 직선이 X 와 Y 의 관계를 근사적으로 가장 잘 표현한다면 이때 X 와 Y 는 선형적인 관계(linear relationship)를 가지고

있다고 한다. 피어슨의 상관계수는 관찰값들이 그들의 선형적인 관계를 가장 잘 설명해주는 직선에 얼마나 가까운지를 나타내주는 척도이다. 피어슨의 상관계수는 다음과 같이 나타나며 $-1 \leq r \leq 1$ 의 값을 가진다.

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

r 의 부호는 한 변수가 증가함에 따라 다른 변수도 증가하는지(+), 감소하는지(-)를 나타내주며, r 값이 1에 가까울수록 점들이 직선에 얼마나 가까운지를 알려준다. 즉, $r=1$ (혹은 -1 이면), 모든 점들이 한 변수가 증가하면 다른 한 변수도 증가하는(혹은 감소하는) 직선상에 있다고 볼 수 있다. 또한 $r=0$ 이면 두 변수는 어떠한 선형 관계도 없다고 본다. 주의할 점은 r 값은 부호가 없으며 두 변수의 인과관계를 설명하는 것은 아니라는 점이다. 또한 두 변수 사이에 비선형 관계가 존재하거나, 한 사람에게서 두 번 이상의 관찰값이 있을 경우, 이상치가 있을 경우, 자료가 부집단들로 구성되어 있으며, 관찰값들의 평균이 다른 경우 등에는 r 값을 계산해서는 안된다.

11.6.2. 스피어만의 순위상관계수(Spearman's rank correlation coefficient)

만일 다음 중 최소한 한 가지에 해당하는 경우라면 피어슨의 상관계수에 대응하는 비모수적인 방법인 스피어만의 순위상관계수를 계산해야 한다.

- 가. X 와 Y 중 최소한 한 변수가 순서형(ordinal)척도로 측정된 경우
- 나. X 와 Y 둘다 정규분포를 하지 않는 경우
- 다. 표본수가 적은 경우
- 라. 두 변수 간의 관계가 선형적이 아닌 경우에 두 변수 간의 관련성을 측

정할 경우

11.7 생존자료분석(Survival data Analysis)

각 개인이 관심 결과 변수(주로 사망 과 같은 사건)에 도달할 때까지 걸린 시간과 관련된 자료분석 방법을 말하며, 크게 다음과 같은 두 가지 특성을 가지고 있다.

가. 주된 관심사가 사건의 여부가 아니라 환자가 사건 발생시점까지 도달하는데 걸린 시간(time to event)이다.

나. 자료가 중도절단(censored) 될 수도 있다.

생존시간(survival time)이란 연구의 시작점 또는 연구대상자의 연구참여 시점(starting point)에서 사건 발생시점까지 도달하는 사이의 시간을 의미한다. 하지만 때로는 연구기간 도중 환자가 사건발생과 무관하게 된 시점만을 알 수 있을 뿐, 실제로 환자가 실제 사건발생에 도달한 시점이 언제인지 알 수 없는 경우도 발생한다. 즉, 연구가 종료되는 시점까지 환자에게서 사건이 발생하지 않거나, 사건이 발생하기 전에 연구 참여자가 참가 동의를 철회하거나 중도탈락 하는 경우, 추적기간이 종료되기 전에 연구와 상관없는 이유로 사망하거나 탈락되는 경우 등이 발생할 수 있다.

11.7.1 생존자료의 표현

가. 생존곡선(survival curve)

기준 시점 이후 모든 시점에 대해 각 연구참여자가 사건발생에 도달하지 않을 수 있는 누적확률(생존확률-survival probability)의 곡선을 생존곡선이라 한다. 즉, 환자가 t 시점까지 생존해 있을 확률의 누적분포로 볼 수 있다.

나. 카플란 마이어(Kaplan-Meier method)

누적 한계 추정법(product-limit method)이란 생존자료에 대한 생존분석을 수행하여 생존함수를 만들어내고 그리는 분석방법이다. 환자 개개인의 생존(혹은 중도 절단) 시간을 관측하여 각 시점에서 생존함수를 추정하게 된다. 종료점이 발생한 경우에만 그 확률이 변하므로, 생존곡선은 계단 모양이 된다. 예를 들어 사망이 D 개의 중복되지 않는 시간 $(t_1 < t_2 < \dots < t_D)$ 에 발생한다고 가정한다. 그 중에서 i 번째 시간에 생존한 n_i 명의 연구참여자 중에서 d_i 명의 연구참여자가 사망했다고 가정한 경우 시간 t 에서의 생존확률 추정치, $S(t)$ 는 다음과 같다.

$$S(t) = \prod_{t_i \leq t} (1 - d_i/n_i). \quad \text{단, } S(t) = 1 \text{ if } t < t_1$$

표 4는 카플란-마이어 방법에서 누적 생존율을 구하는 한 예시를 보여준다.

표 4 카플란-마이어 방법의 예시

t_i	d_i	n_i	$S(t)$
6	3	21	$(1-3/21)=0.857$
7	1	17	$0.857 \times (1-1/17)=0.807$
10	1	15	$0.807 \times (1-1/15)=0.753$
13	1	22	$0.753 \times (1-1/12)=0.690$
16	1	11	$0.690 \times (1-1/11)=0.628$
22	1	7	$0.628 \times (1-1/7)=0.538$
23	1	6	$0.538 \times (1-1/6)=0.448$

11.7.2 생존자료의 표현 생존곡선의 비교

가. 로그-순위 검정 (log-rank test)

로그-순위 검정은 집단들 간의 생존 시간에 유의한 차이가 없다는 귀무가설을 검정하기 위한 비모수적 방법이다. 각 시점에 대해, 그룹1 과 2에 대하여 각각 사망-생존 을 표시한 2×2 분할표를 표 5와 같이 작성할 수 있다.

표 5 두 군 사이의 사망수와 생존수를 표현한 2×2 분할표

	사망	생존	계
그룹1	d_{1i}	$n_{1i}-d_{1i}$	n_{1i}
그룹2	d_{2i}	$n_{2i}-d_{2i}$	n_{2i}
	d_i	n_i-d_i	n_i

※ $i=1, \dots, D$ 는 각 군에서 D 개의 중복되지 않은 시간임.

n_{1i}, n_{2i}, d_i 가 고정되었을 때, d_{1i} 의 조건부 분포는 초기하 분포 (hypergeometric distribution)이다. 그러므로 d_i 의 기댓값과 분산은 다음과 같다.

$$E(d_i) = E_{1i} = \frac{n_{1i}d_i}{n_i}$$

$$V(d_i) = V_{1i} = \frac{d_i(n_{1i}/n_i)(1-n_{1i}/n_i)(n_i-d_i)}{n_i-1}$$

위의 기댓값과 분산을 이용해 다음과 같은 검정 통계량 값을 구한다.

$$Z = \frac{\sum_{i=1}^D (d_{1i} - E_{1i})}{\sqrt{\sum_{i=1}^D V_{1i}}} .$$

검정 통계량 Z 는 근사적으로 표준정규분포를 따르므로 그에 따른 검정을 실시한다.

나. 회귀모형-콕스의 비례위험모형(Cox's proportional hazard(PH) model)

$$h_i(t) = h_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki})$$

여기서 $h_i(t)$: i 번째 개인의 t 시점의 위험

즉, 여러 가지 변수들로 i 번째 사람의 t 시점의 위험율(hazard rate)을 구하는 회귀방정식이다. 각 대상자에 대해 연구종료 시점에 이르는 기간을 예측하는데 있어서 회귀분석을 사용한다. 대상자들의 특성차이 뿐만 아니라, 차이가 있는 관찰기간까지 통제 가능하다. 이들 추정치의 지수값, 즉, 변수들의 $\exp(\beta_k)$ 값들은 상대위험 추정치(relative hazard) 또는 위험비(hazard ratio)라고 하며, 각 변수가 한 단위 증가할 때 질병발생의 추정위험이 상대적으로 얼마나 높은지 알려주는 값이다.

이 모형에서는 전체 연구기간 동안 상대위험이 일정하게 유지된다는 가정이 필수이다. 비교하고자 하는 집단들 간의 위험은 서로 일정하게 비례한다는 의미에서 비례위험 모형 이라고 하기도 한다. 이 비례위험의 가정은 생존곡선을 그려서 확인해 보는 방법이 있으며 모형 내에 log(시간) 변수와 공변량 간의 상호작용 항을 추가한 뒤 상호작용 항이 유의한 지 검정하여 조사하는 방법 등이 있다.

11.8 생분석에서의 가설 검정

11.8.1. 일치성 검정 (Test for Equality)

$\varepsilon = \lambda_1 - \lambda_2$ 를 대조군과 실험군의 위험률(hazard rate)의 차이라고 하면, 위험률의 차이를 검정하기 위한 가설은 다음과 같다.

$$H_0: \varepsilon = 0 \quad \text{vs.} \quad H_a: \varepsilon \neq 0$$

i 번째 처리그룹의 위험률 λ_i 의 최우추정량(MLE) $\hat{\lambda}_i$ 은 평균이 λ_i 이고 분산이 $\sigma^2(\lambda_i)$ 인 정규분포를 따르므로, $n_1 = kn_2$ 라 가정하면, 유의수준 α 와 검정력이 $1 - \beta$ 일 경우 일치성 검정을 위한 표본수는 다음과 같이 산출된다.

$$n_2 = \frac{(z_{\alpha/2} + z_{\beta})^2}{(\lambda_1 - \lambda_2)^2} \left[\frac{\sigma^2(\lambda_1)}{k} + \sigma^2(\lambda_2) \right]$$

여기서 $z_{\alpha/2}$ 를 사용한 것은 가설이 양측검정이기 때문이며, z_{β} 를 사용한 것은 검정력이 $(1 - \beta)100\%$ 보다 커야 하므로 단측검정이 적용되기 때문이다.

11.8.2. 비열등성/우월성 검정(Test for Non-Inferiority/Superiority)

일반적으로 실험군의 위험률이 더 낮는데 관심을 가지게 된다. 따라서 $\varepsilon = \lambda_1 - \lambda_2$ 이므로 $\varepsilon < 0$ 인 경우 실험군의 performance가 대조군에 비해 더 좋다고 할 수 있다. 따라서 비열등성과 우월성 검정에서의 가설은 다음과 같이 설정될 수 있다.

$$H_0: \varepsilon \leq \delta \quad \text{vs.} \quad H_a: \varepsilon > \delta$$

여기서 δ 는 우월성 또는 비열등성의 한계(margin)로, 만약 $\delta > 0$ 인 경우에 귀무가설의 기각은 실험군이 대조군에 비해 우월(superiority)하다는 것을 제시하며, $\delta < 0$ 인 경우에 귀무가설의 기각은 실험군이 대조군에 비해 비열등(non-inferiority)하다는 것을 의미한다. 일치성 검정에서와 비슷하게

$n_1 = kn_2$ 라 가정하면, 유의수준 α 와 검정력이 $1 - \beta$ 일 경우 비열등성/우월성 검정을 위한 표본수는 다음과 같이 산출된다.

$$n_2 = \frac{(z_\alpha + z_\beta)^2}{(\varepsilon - \delta)^2} \left[\frac{\sigma^2(\lambda_1)}{k} + \sigma^2(\lambda_2) \right]$$

여기서 z_α 를 사용한 것은 가설이 비열등성/우월성 검정은 단측검정이기 때문이며, z_β 를 사용한 것은 검정력이 $(1 - \beta)100\%$ 보다 커야 하므로 역시 단측검정이 적용되기 때문이다.

11.8.3. 동등성 검정(Test for Equivalence)

동등성 검정의 가설은 다음과 같다.

$$H_0: |\varepsilon| \geq \delta \quad \text{vs.} \quad H_a: |\varepsilon| < \delta$$

따라서 $n_1 = kn_2$ 라 가정하면, 유의수준 α 와 검정력이 $1 - \beta$ 일 경우 동등성 검정을 위한 표본수는 다음과 같이 산출된다.

$$n_2 = \frac{(z_\alpha + z_{\beta/2})^2}{(\delta - |\varepsilon|)^2} \left[\frac{\sigma^2(\lambda_1)}{k} + \sigma^2(\lambda_2) \right]$$

통계적 검정에서 제 1종 오류는 귀무가설이 참인데도 귀무가설을 잘못 기각하는 오류이므로 동등성 검정에서 제 1종 오류는 한쪽만 생각하면 된다. 그러나 제 2종 오류는 귀무가설이 거짓인데 귀무가설을 잘못 채택할 때 발생한다. 따라서 동등성 검정에서 제 2종 오류는 귀무가설을 잘못 받아들이는 경우는 ε 의 부호에 따라 달라지므로 양측을 생각하여야 한다. 이와 같은 이유로 표본크기 계산식에서 z_α 와 $z_{\beta/2}$ 가 사용되었다.

11.8.4. 예제

연구의 후원자(sponsor)는 Hodgkin's disease(HOD) 또는 non-Hodgkin's lymphoma(NHL)를 가진 환자들을 대상으로 임상시험을 하고자 한다. 이들 환

자들은 고농도의 화학요법 적용 후 HLA-matched sibling donor로부터 allogeneic(allo) 골수이식을 받거나 또는 autogeneic(auto) 골수이식을 받게 된다. 이 연구의 목적은 allo 또는 auto 골수이식을 받은 환자들 간에 백혈병의 생존시간을 비교하고자 한다. 이 시험은 1년의 accrual($T_0 = 1$)을 가지고 3년간($T = 3$) 지속하고자 한다. allo와 auto 골수이식 군의 환자는 일정하다고(uniform) 가정하고($\gamma = 0$), allo 군과 auto 군의 leukemia-free 위험률은 각각 1($\lambda_1 = 1$)과 2($\lambda_2 = 2$)로 가정한다. 따라서 두 군의 위험률에 대한 분산은 다음과 같이 계산되어 진다.

$$\sigma^2(\lambda_i) = \lambda_i^2 \left(1 + \frac{e^{-\lambda_i T} - e^{-\lambda_i(T+T_0)}}{\lambda_i T_0} \right)^{-1}$$

11.8.4.1 일치성 검정 (Test for equality)

두 처리그룹간의 표본크기가 동일하다면($n = n_1 = n_2$), 일치성 검정의 표본크기 계산과정은 다음과 같다.

- ① 가설: $H_0: \varepsilon = \lambda_1 - \lambda_2 = 0$ vs $H_a: \varepsilon = \lambda_1 - \lambda_2 \neq 0$.
- ② 유의수준 $\alpha = 0.05$
- ③ Power $(1 - \beta)100\% = 80\%$. 즉, 제 2종 오류 $\beta = 0.2$
- ④ 이상의 조건에서 표본크기는 다음과 같다.

$$\begin{aligned} n &= \frac{(z_{\alpha/2} + z_{\beta})^2}{(\lambda_2 - \lambda_1)^2} \left[\frac{\sigma^2(\lambda_1)}{k} + \sigma^2(\lambda_2) \right] \\ &= \frac{(1.96 + 0.84)^2}{(2 - 1)^2} (0.97 + 3.94) \\ &\approx 39 \end{aligned}$$

즉, 두 그룹의 표본크기는 각각 39명이여야 한다.

11.8.4.2 우월성 검정 (Test for superiority)

두 처리그룹간의 표본크기가 동일하고($n = n_1 = n_2$), 우월성의 한계(margin) $\delta = 0.2$ 로 가정하면, 우월성 검정의 표본크기 계산과정은 다음과 같다.

- ① 가설: $H_0: \varepsilon \leq \delta$ vs. $H_a: \varepsilon > \delta$
- ② 유의수준 $\alpha = 0.05$
- ③ Power $(1 - \beta)100\% = 80\%$. 즉, 제2종 오류 $\beta = 0.2$
- ④ 우월성의 한계 $\delta = 0.2$
- ⑤ 이상의 조건에서 표본크기는 다음과 같다.

$$\begin{aligned} n &= \frac{(z_\alpha + z_\beta)^2}{(\lambda_2 - \lambda_1 - \delta)^2} \left[\frac{\sigma^2(\lambda_1)}{k} + \sigma^2(\lambda_2) \right] \\ &= \frac{(1.64 + 0.84)^2}{(2 - 1 - 0.2)^2} (0.97 + 3.94) \\ &\approx 48 \end{aligned}$$

즉, 두 그룹의 표본크기는 각각 48명이어야 한다.

11.8.4.3 동등성 검정 (Test for equivalence)

두 처리그룹간의 표본크기가 동일하고($n = n_1 = n_2$), $\lambda_1 = \lambda_2 = 1$ 이며, 동등성의 한계(margin) $\delta = 0.5$ 로 가정하면, 우월성 검정의 표본크기 계산과정은 다음과 같다.

- ① 가설: $H_0: |\varepsilon| \geq \delta$ vs. $H_a: |\varepsilon| < \delta$
- ② 유의수준 $\alpha = 0.05$
- ③ Power $(1 - \beta)100\% = 80\%$. 즉, 제2종 오류 $\beta = 0.2$
- ④ 동등성 한계 $\delta = 0.5$
- ⑤ 이상의 조건에서 표본크기는 다음과 같다.

$$\begin{aligned} n &= \frac{(z_\alpha + z_{\beta/2})^2}{\delta^2} \left[\frac{\sigma^2(\lambda_1)}{k} + \sigma^2(\lambda_2) \right] \\ &= \frac{(1.64 + 1.28)^2}{(0.5 - 0)^2} (0.97 + 0.97) \\ &\approx 67 \end{aligned}$$

즉, 두 그룹의 표본크기는 각각 67명이다.

[붙임]

산·학·연 전문가 협의체 위원

- | | | | |
|---|----------------|-----|-----------------------------|
| ○ | 고려대학교산학협력단 | 안형진 | ※ 동 가이드라인 관련
연구사업 수행 책임자 |
| ○ | 고려대학교산학협력단 | 이정복 | |
| ○ | 동국대학교 | 주용성 | |
| ○ | 연세대학교 보건대학원 | 강대룡 | |
| ○ | 서울아산병원 의학통계학교실 | 윤성철 | |

의료기기 임상시험 관련 통계기법 가이드라인

발행처	식품의약품안전평가원
발행일	2010년 8월
발행인	김승희
편집위원장	김동섭, 유규하
편집위원	오현진, 조양하, 남봉현, 오현주, 이병영, 이원규, 홍충만, 장정훈, 김별아, 방지영, 이정연, 우대곤, 임경미, 배우진 122-704 서울시 은평구 통일로 194 식품의약품안전평가원 의료기기연구과
문의처	식품의약품안전청 재료용품과 식품의약품안전청 통상통계담당관

전화: 02-380-1754~56 / 02-350-4452~68 / 02-380-1661~2

팩스: 02-351-3726 / 02-350-4470 / 02-380-2893