

소기업·연구개발자를 위한 알기 쉬운 제품화 가이드

- 의료제품 개발단계의 쉬운 임상통계 -



지침서·안내서 제·개정 점검표

명칭	소기업·연구개발자를 위한 알기 쉬운 제품화 가이드 - 의료제품 개발단계의 쉬운 임상통계-
----	--

아래에 해당하는 사항에 체크하여 주시기 바랍니다.

등록대상 여부	☐ 이미 등록된 지침서·안내서 중 동일·유사한 내용의 지침서·안내서가 있습니까?	☐ 예 ☒ 아니오
	☒ 상기 질문에 '예'라고 답하신 경우 기존의 지침서·안내서의 개정을 우선적으로 고려하시기 바랍니다. 그럼에도 불구하고 동 지침서·안내서의 제정이 필요한 경우 그 사유를 아래에 기재해 주시기 바랍니다. (사유 :)	
	☐ 법령(법·시행령·시행규칙) 또는 행정규칙(고시·훈령·예규)의 내용을 단순 편집 또는 나열한 것입니까?	☐ 예 ☒ 아니오
	☐ 단순한 사실을 대외적으로 알리는 공고의 내용입니까?	☐ 예 ☒ 아니오
	☐ 일회성 지시·명령에 해당하는 내용입니까?	☐ 예 ☒ 아니오
	☐ 외국 규정을 단순 번역하거나 설명하는 내용입니까?	☐ 예 ☒ 아니오
	☐ 신규 직원 교육을 위해 법령 또는 행정규칙을 알기 쉽게 정리한 자료입니까?	☐ 예 ☒ 아니오
	☒ 상기 사항 중 어느 하나라도 '예'에 해당되는 경우에 지침서·안내서 등록 대상이 아닙니다. 지침서·안내서 제·개정 절차를 적용하실 필요는 없습니다.	
지침서·안내서 구분	☐ 행정사무의 통일을 기하기 위하여 내부적으로 행정사무의 세부기준이나 절차를 제시하는 것입니까? (공무원용)	☐ 예(☒ 지침서) ☒ 아니오
	☐ 민원인들의 이해를 돕기 위하여 법령 또는 행정규칙을 알기 쉽게 설명하거나 특정 민원업무에 대한 행정기관의 대외적인 입장을 기술하는 것입니까? (민원인용)	☒ 예(☒ 안내서) ☐ 아니오
기타 확인 사항	☐ 상위 법령을 일탈하여 새로운 규제를 신설·강화하거나 민원인을 구속하는 내용이 있습니까?	☐ 예 ☒ 아니오
	☒ 상기 질문에 '예'라고 답하신 경우 상위법령 일탈 내용을 삭제하시고 지침서·안내서 제·개정 절차를 진행하시기 바랍니다.	

상기 사항에 대하여 확인하였음.

2026 년 9 월 일

담당자

박 진 경

확 인(부서장)

김 희 성

이 안내서는 소기업과 연구개발자들을 대상으로 의료제품 제품화를 위한 임상시험 설계 시 통계적 고려사항에 대하여 기본적으로 알아야 할 개념들을 알기 쉽게 설명하거나 식품의약품안전처의 입장을 기술한 것입니다.

본 안내서는 대외적으로 법적 효력을 가지는 것이 아니므로 본문의 기술방식(“~하여야 한다” 등)에도 불구하고 참고로만 활용하시기 바랍니다. 또한, 이 안내서는 2026년 9월 현재의 과학적·기술적 사실 및 유효한 법규를 토대로 작성되었으므로 이후 최신 개정 법규 내용 및 구체적인 사실관계 등에 따라 달리 적용될 수 있음을 알려드립니다.

※ “민원인 안내서”란 민원인들의 이해를 돕기 위하여 법령 또는 행정규칙을 알기 쉽게 설명하거나 특정 민원업무에 대한 행정기관의 대외적인 입장을 기술하는 것(식품의약품안전처 지침서등의 관리에 관한 규정 제2조)

※ 본 안내서에 대한 의견이나 문의사항이 있을 경우 사전상담과에 문의하시기 바랍니다.

전화번호 : 043-719-2935

팩스번호 : 043-719-2910

제·개정 이력

연번	제·개정번호	승인일자	주요내용
1	안내서-1530-01	2026.09.	제정
2			

한눈에 보는 임상시험 설계 핵심요소

임상시험의 신뢰성과 타당성을 확보하기 위해서는 연구설계 단계에서 반드시 고려해야 할 핵심요소들이 있습니다. 본 안내서에서는 임상시험 설계 시 중요한 다섯가지 통계학적 요소를 중심으로 기본 개념과 적용방법을 설명합니다. 아래의 요약표를 통해 전체 내용을 먼저 살펴본 후 본문을 읽으면 각 개념을 보다 쉽게 이해할 수 있습니다.

핵심요소	주요내용
평가변수	임상시험에서 무엇을 평가할 것인지 미리 정하는 기준으로, 임상시험의 성공 여부를 판단하는 핵심지표입니다.
대상자수	신뢰할 수 있는 결과를 얻기 위해 필요한 시험대상자 수를 결정하는 과정입니다.
대조군 선정	임상시험 결과를 객관적으로 비교할 수 있도록 적절한 비교대상(대조군)을 선정합니다.
무작위배정과 눈가림	비뚤림(편향)을 줄이고, 공정한 비교가 이루어지도록 임상시험대상자를 시험군과 대조군에 무작위로 배정하고 해당 정보를 가려 연구 결과의 객관성과 신뢰성을 높이는 방법입니다.
결측치 및 중도탈락	자료 누락이나 중도 탈락이 발생하더라도 결과의 신뢰성을 유지할 수 있도록 사전에 계획하고 관리합니다.

임상시험 설계 핵심요소 요약

목 차

제 1 장 | 의료제품 개발과 통계의 이해

- 01 1.1 왜 통계가 필요한가요
- 02 1.2 의료제품개발 과정에서의 통계 활용
- 03 1.3 임상시험 설계 시 통계적 핵심 요소

제 2 장 | 주요 통계 실무 고려사항

- 04 2.1 평가변수
- 07 2.2 대상자수
- 10 2.3 대조군 선정
- 11 2.4 무작위배정과 눈가림
- 13 2.5 결측치 및 중도탈락
- 14 2.6 [의약품 예시] 본태성 고혈압 환자 대상의 유효성 평가
- 18 2.7 [의료기기 예시] CT영상 기반 질환위험도 분석에 대한 유효성 평가

제 3 장 | 소기업과 연구개발자들을 위한 준비사항

- 22 3.1 CRO 활용 시 확인 사항
- 22 3.2 임상시험 설계 단계 권고사항
- 24 3.3 규제기관 활용

- 25 부록 1. 주요 통계 가이드라인
- 25 부록 2. 평가변수별 주요 통계 분석법
- 26 부록 3. 주요 용어
- 32 찾아보기



1.1 | 왜 통계가 필요한가요

의약품, 바이오의약품 및 의료기기¹⁾는 개발 과정에서 의료제품²⁾의 효과, 성능 및 안전성을 과학적이고 객관적으로 입증해야 합니다. 의료제품이 실제로 기대한 효과를 나타내는지, 기존 제품과 비교하여 어떠한 차이가 있는지, 그리고 사용 시 안전한지를 평가하기 위해서 다양한 자료를 수집하고 분석하는 과정이 필요합니다. 통계는 이러한 자료를 객관적으로 해석하고 의사결정을 지원하는 도구입니다. 단순히 임상시험³⁾이 종료된 후 결과를 분석하는데 사용하는 것이 아니라, 개발 초기에서부터 평가계획을 수립하고 임상시험을 설계하는 과정 전반에 활용됩니다.

의료제품 개발을 시작할 때에는 무엇을 평가할 것인지(평가변수)를 정해야 하며, 신뢰할 수 있는 결과를 얻기 위해 필요한 대상자수/표본수를 검토해야 합니다. 또한 의료제품의 유효성 및 안전성을 객관적으로 확인하기 위해 적절한 대조군을 선정하고, 임상시험 과정에서 발생할 수 있는 편향(bias)을 최소화하기 위한 방법을 고려해야 합니다. 임상시험이 진행되는 동안에는 결측치나 중도탈락과 같은 문제를 관리하여 결과의 신뢰성을 확보할 필요가 있습니다.

특히, 중소기업의 경우 연구개발, 유효성/성능평가, 인허가 업무를 제한적 인력으로 수행하는 경우가 많습니다. 이때 통계에 대한 기본적인 이해는 외부 전문가와 원활한 소통뿐만 아니라 임상시험 계획 수립, 자료 검토 및 결과 해석 과정에서도 중요한 역할을 합니다.

본 안내서는 통계 전문가가 아닌 연구개발자 혹은 인허가 담당자를 대상으로 의료제품 개발 과정에서 자주 접하게 되는 기본적인 통계 개념을 소개하고, 이를 실제 사례에 적용할 수 있도록 이해를 돕기 위해 마련되었습니다. 이를 통해 의료제품 개발 초기 단계부터 통계적 고려사항을 반영하고, 보다 체계적인 제품화 전략을 수립하는 데 도움을 주고자 합니다.

예를 들어, 한 기업이 새로운 의료제품의 유효성/성능을 확인하기 위해 임상시험을 수행했을 때, 임상시험 결과에서는 의료제품 사용 후 개선 경향이 관찰되었지만, 개발 초기 단계에서 평가변수와 대상자수를 충분히 검토하지 않아 결과를 명확하게 해석하기 어려웠다고 하는 경우가 있을 수 있습니다.

1) 의료기기, 체외진단의료기기를 통칭하여 용어 사용하였습니다(이하 “의료기기”).

2) 의약품, 바이오의약품 및 의료기기(체외진단의료기기포함)를 통칭하여 의료제품이라 표기하였습니다.

3) 의료기기의 경우, 의료기기 임상시험, 체외진단의료기기 임상적 성능시험을 통칭하여 용어 사용하였습니다(이하 “임상시험”).

이에 따라, 추가 시험이 필요하게 되었고, 개발 일정이 지연되면서 예상보다 많은 시간과 비용이 소요될 것입니다. 반면, 어떤 기업은 개발 초기부터 평가변수, 대상자수 및 대조군을 사전에 검토하여 임상시험을 설계하였고, 확보된 자료를 의료제품 개발 및 허가자료로 활용할 수 있었습니다. 이처럼 통계는 임상시험 결과를 분석하기 위한 도구가 아닌, 의료제품 개발 과정의 효율성과 성공 가능성에 미치는 중요한 요소입니다(그림 1).



그림 1. 통계적 고려 여부에 따른 의료제품 개발 과정 예시

1.2 | 의료제품 개발과정에서의 통계 활용

의약품, 바이오의약품 및 의료기기의 개발 과정은 의료제품의 임상시험 설계에서부터 임상시험 수행, 결과 분석 및 허가에 이르기까지 여러 단계로 이루어집니다. 각 단계는 의료제품의 유효성/성능 및 안전성을 객관적으로 평가하기 위한 다양한 의사결정이 이루어지며, 이 과정에서 통계적 고려사항이 활용됩니다.

예를 들어, 임상시험계획 수립단계에서는 무엇을 측정할 것인지(평가변수), 몇 명을 평가할 것인지(대상자수), 무엇과 비교하여(대조군) 유효성/성능을 입증할지 결정해야 합니다. 또한 임상시험 수행단계에서는 무작위배정과 눈가림을 통해 편향을 최소화하고, 자료 수집과정에서 결측치와 중도탈락에 대한 관리 또한 필요합니다. 이와 같이 통계는 특정 단계에 국한되지 않고 제품개발 전과정에서 객관적인 근거를 제공하는 역할을 수행합니다(그림 2).

임상시험에서의 통계는 임상시험 중에서 발생할 수 있는 불확실성을 정량적으로 평가하고 관찰된 결과가 우연에 의한 것인지 실제 효과에 의한 것인지를 판단하기 위한 도구로 사용되기에, 임상시험 결과를 근거 기반의 의사결정으로 연결하는 핵심요소이며, 허가 심사 근거확보를 위한 필수조건이라고 할 수 있습니다.

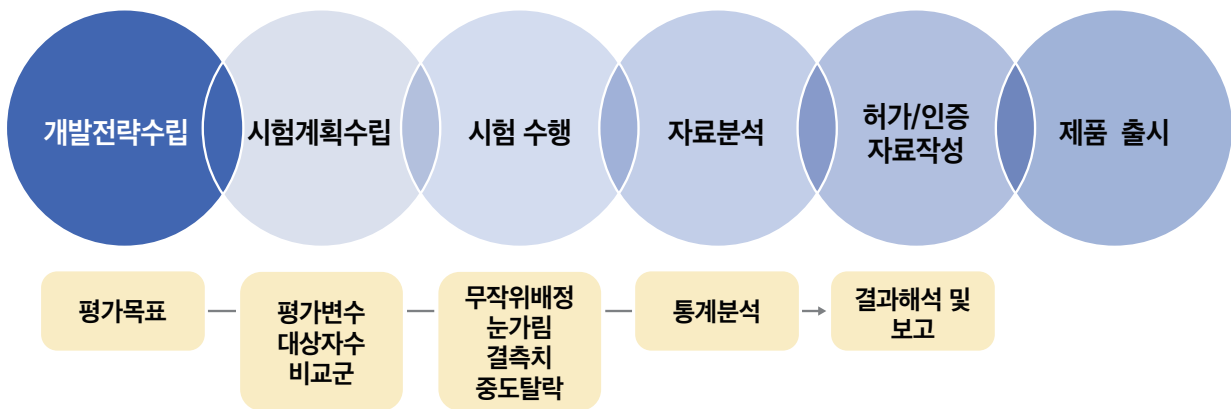


그림 2. 의료제품 개발과정에서의 통계 활용

1.3 | 임상시험 설계 시 통계적 핵심 요소

의료제품개발 과정에서 수행하는 임상시험의 목적은 (의약품의) 유효성 혹은 (의료기기의) 성능 및 안전성을 객관적으로 확인하는 데 있습니다. 이 과정에서 다양한 통계적 고려사항이 요구되어지며, 특히 임상시험 설계와 자료의 신뢰성 확보에 직접적인 영향을 미치는 핵심 요소들은 아래와 같습니다(그림 3).



그림 3. 임상시험 설계 시 통계적 핵심요소



임상시험 설계단계에서는 임상시험 목적과 결과 해석의 기준이 되는 통계적 요소들을 사전에 결정해야 합니다. 이 결정들은 임상시험의 타당성, 결과의 신뢰성 및 규제적 수용성에 직접적인 영향을 미칠 수 있습니다. 본 장에서는 앞 절에서 언급된, 임상시험 설계 시 실무자가 반드시 고려해야 하는 통계적 핵심요소 개념을 설명하고자 합니다⁴⁾.

2.1 | 평가변수

평가변수(Endpoint)는 임상시험에서 유효성/성능 또는 안전성을 측정하기 위해 사용되는 항목으로, 임상시험을 통해 “무엇을 평가할 것인가”를 구체적으로 나타내는 지표입니다. 평가변수는 임상시험의 목적을 고려하여 설정되어야 하며, 결과 해석 및 허가자료의 근거로 활용되므로 명확하고 객관적으로 정의되어야 합니다.

평가변수는 일차평가변수 (Primary endpoint)와 이차평가변수 (Secondary endpoint), 탐색평가변수 (Exploratory endpoint), 안전성평가변수 (Safety endpoint) 등으로 나뉩니다. 일차평가변수는 임상시험의 일차목적(Primary objective)을 평가하기 위한 핵심지표로, 이를 충족하여야 임상시험의 “성공”여부를 뒷받침하는 첫 단계가 됩니다. 이차평가변수는 일차목적과 관련된 보조적인 측정지표이거나 이차적 목적과 관련된 효과 지표입니다.

복수의 일차평가변수를 설정하는 경우, 왜 각각의 평가변수가 중요한지를 과학적, 임상적으로 설명해야 할 수 있습니다. 또한, 결과 해석이 복잡해지고 통계적 해석의 신뢰성이 낮아질 수 있으므로, 임상시험 주요목적(Primary objective)을 명확히 반영하고 결과의 해석 가능성을 높일 수 있도록 설정하는 것을 권장합니다. 일반적으로 임상시험 대상자수(혹은 표본수)를 추정하는 데 일차평가변수를 이용합니다.

예를 들어, 본태성 고혈압 환자 치료제 임상시험의 경우, 유효성을 평가하기 위해 기저 대비 임상시험용 의약품 투여 8주 시점의 MSSBP 변화량을 일차평가변수로 설정하여 대조군과 비교하는 임상시험을 설계할 수 있습니다. 그 외, 이차평가변수로 기저대비 4주 시점의 MSSBP 변화량이라던가 기저대비 4주, 8주 맥압의 변화량 등을 설정할 수 있고, 해당 시점의 혈압 정상화 비율도 이차평가변수로 설정할 수 있습니다. 안전성 평가변수로는

⁴⁾ 설계 시 고려되는 통계개념이 아니지만, 결과 해석 시 매우 중요한 통계적 개념인 유의확률(p-value)과 신뢰구간(CI, confidence interval)은 반드시 알아야 하는 통계개념으로, 자세한 내용은 [부록 3. 주요용어]를 참고하시기 바랍니다.

이상사례, 활력징후, 실험실적 검사, 심전도 검사 등을 통해 이상사례 중증도나 임상시험용 의약품과의 인과관계를 평가합니다.

의료기기 임상시험의 경우에는 예를 들어, 특정질환 대상자를 선별하는 소프트웨어의 경우, AUROC (Area Under the Receiver Operating Characteristic curve), 민감도와 특이도, 정확도 등이 평가변수가 될 수 있습니다. 그 외 민감도와 특이도, 정확도 등은 이차평가변수로 설정할 수 있습니다.

평가변수는 사전에 명확히 정의되어야 하며, 정의되지 않은 지표를 사후적으로 분석하여 주요 결과로 사용하는 것은 일반적으로 인정되지 않습니다. 따라서, 임상시험 계획 단계에서 평가변수를 명확히 설정하는 것은 매우 중요합니다. 또한, 해당 지표가 객관적으로 측정 가능하고, 반복측정시에도 유사한 결과를 얻을 수 있는 재현성 확보도 중요한 판단 기준이 될 수 있습니다. 이때 중요한 기준은 규제와 임상시험 가이드라인의 부합성입니다. 이는 해당 평가변수가 임상적 유효성이나 성능을 허가 및 심사 과정에서 인정 가능한 방식으로 입증할 수 있는지를 의미합니다.

평가변수는 단순히 측정항목이 아니라 임상시험의 목적과 결론을 연결하는 핵심 기준이므로, 임상시험 설계 단계에서 충분한 검토와 합의를 통해 결정되어야 합니다.

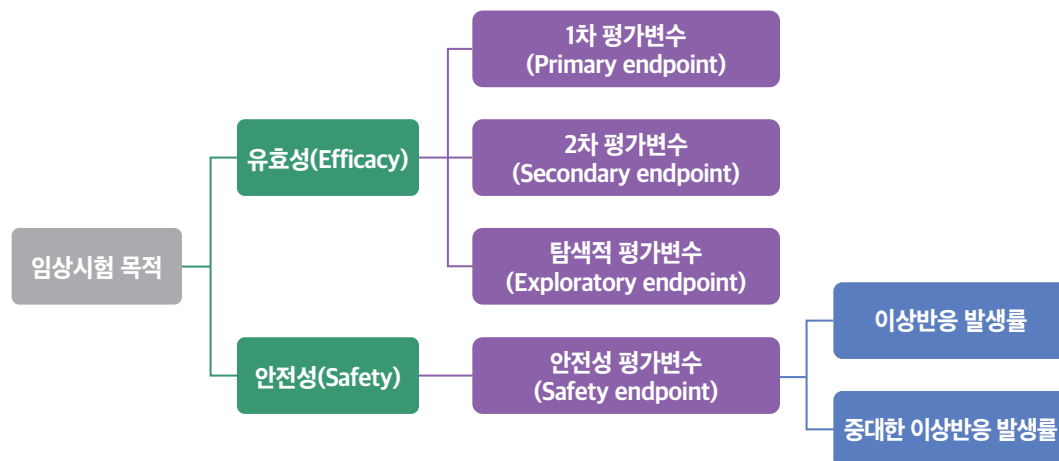


그림 4. 평가변수 체계

의료제품	제품분류	일차평가변수 예시	이차평가변수 예시
의료기기	체외진단 의료기기	임상적 민감도 및 임상적 특이도	■ 양성예측도, 음성예측도, 우도비 ■ 기허가 제품과 상관성 ■ 양성일치율, 음성일치율, 총일치율, kappa 값 등
	디지털 의료기기	기저대비 공황장애 심각도 척도(PDSS) 변화량	■ 반응률 ■ 관해율 ■ 전반적 기능 개선율 등
의약품	고혈압 치료제	기저 대비 임상시험용 의약품 투여 후 8주 시점의 MSSBP 변화량	■ 기저 대비 임상시험용 의약품 투여 후 4주 시점 MSSBP 변화량 ■ 기저 대비 임상시험용 의약품 투여 후 4주, 8주 시점의 혈압 정상화 비율 및 혈압 반응률 등
	당뇨병 치료제	기저 대비 임상시험용 의약품 투여 후 24주 시점의 HbA1c 변화량	■ 기저 대비 임상시험용 의약품 투여 후 6, 12, 18주 시점의 HbA1c 변화량 ■ 기저 대비 임상시험용 의약품 투여 후 6, 12, 18, 24 주 시점의 FPG 변화량 등

표 1. 임상시험 시 유형별 평가변수 예시

평가변수를 설정할 때에는 임상적 타당성, 재현가능성, 가이드라인 부합성 등을 고려하여 유효성 평가 변수와 안전성 평가변수 등을 정의하여야 하고, 임상시험 목적에 맞는 일차평가변수 정의가 임상시험 성공여부를 판가름하는 설계의 첫 단추라 할 수 있습니다.

2.2 | 대상자수

대상자수(Sample size)는 임상시험에 포함되는 대상자수/표본수⁵⁾를 의미합니다. 임상시험에서 유효성/성과 안전성을 신뢰성 있게 평가하기 위해 적절한 대상자수를 설정하는 것이 중요합니다. 대상자수가 지나치게 적은 경우, 실제 효과나 성능 차이가 존재하더라도 통계적으로는 확인되지 않아 분석결과가 실패로 나올 수 있고, 반대로 불필요하게 많은 대상자를 모집하면 개발 시간과 비용이 증가하여 임상시험의 효율이 떨어질 뿐 아니라 윤리적 문제가 발생할 수 있습니다.

대상자수 산정은 임상시험에서 필요한 대상자수를 사전에 결정하는 과정으로 임상시험 결과의 신뢰성과 타당성 확보를 위해 반드시 수행되어야 합니다. 대상자수 산정은 임상시험의 과학적 타당성과 윤리성을 동시에 고려하여 적절한 수준으로 결정되어야 합니다. 즉, 임상시험의 목적과 일차평가변수의 특성, 예상되는 효과크기 등을 고려하여 적절한 대상자수를 산정할 필요가 있습니다. 또한, 임상시험 진행 중 중도탈락이나 결측자료(혹은 결측치)가 발생할 가능성을 고려하여 추가 대상자수를 반영하여 모집하여야 합니다.

예를 들어, 의료기기 임상시험⁶⁾에서 AUROC 의 경우, 예상성능은 기존 시험기기 연구를 통해 0.88임을 확인하였다고 가정합니다. 또한, 성공기준치는 문헌 등을 통해 0.8로 설정하였다고 가정해봅시다. 대상자수 산정에 필요한 통계적 개념이 추가적으로 필요한데, 검정력과 유의수준이 이에 해당됩니다. 즉, 대상자수 산정을 위해 예상 효과크기(예상성능, 0.88)과 성공기준치(참고치, 0.80), 그리고 검정력과 유의수준, 경우에 따라 중도탈락률 등을 고려할 수 있습니다. 대상자수 산정방법 및 근거는 임상시험 계획서에 사전에 정의하는 것이 바람직합니다.

여기서, 통계적 가정인 유의수준은 통상적으로 0.05(즉, 5%)로 설정합니다. 진실은 효과가 없지만 효과가 있다고 결론을 잘못 내릴 확률을 5% 미만으로 허용하겠다는 의미입니다. 자료 분석을 통해 추정된, ‘효과가 없을 때 효과가 있다고 할 확률’이 5%가 넘는다면, 효과가 없다고 할 수 있다고 하고, 반대로 5% 미만으로 계산된다면, 효과가 없다고 할 수 없다는 통계적 판단을 하겠다는 의미입니다. 유의수준 0.05는 일반적으로 양측을 의미합니다. 효과크기의 방향은 설정할 수 없지만 큰 방향이던, 작은 방향이던 참고 유효성/성능효과와의 통계적으로 차이가 있는지를 판단하고자 할 때 양측을 사용합니다. 하지만, 위 예에서는 유사 연구들의 평균 성능 “이상”임(우월성)을 검증하고자 하였으므로 단측 유의수준 0.025로 설정하는 것과 동일한 의미입니다.

5) 임상시험의 분류에 따라 대상자수 혹은 표본수로 나타낼 수 있으나, 대상자수로 통일하여 용어 사용하였습니다(이하 “대상자수”).

6) 의약품의 경우는 2.6절 사례를 참고하시기 바랍니다.

그리고 검정력의 경우, 일반적으로 0.8 (즉, 80%) 이상으로 가정합니다⁷⁾. 예를 들어, 좀 더 강한 검정력 (즉, 효과가 있을 때 통계적으로도 효과가 있다고 결론내릴 확률) 인 0.9로 가정할 수 있습니다. 그리고, 중도탈락율(예를 들어, 20%) 가정하게 되면 최종 대상자수를 산정할 수 있습니다. 대상자수 산정을 위해 필요한 가정과 그 예시 값들을 정리하면 [표 2]와 같습니다.

가정	가정 값
목표 성능효과 (AUROC)	0.88
참고 성능효과 (기준 AUROC)	0.80
검정력	0.9 (90%)
유의수준	0.025 (단측 2.5%)
중도탈락률	0.2 (20%)

표 2. 대상자수 산출 시 필요한 가정

대상자수 산출을 위해서는 통계적 확률분포를 이용하여 수식으로 산출할 수도 있으나, 대상자수 산출 전용 소프트웨어를 이용하여 해당 결과를 [그림 5]와 같이 캡처하여 임상시험계획서에 첨부할 수도 있습니다. [표 3]은 [표 2]의 가정 값들과 그에 해당하는 프로그램 [그림 5]에서의 표현을 매칭시켜 간략하게 나타내었습니다.

Numeric Results											
Solve For:		Sample Size									
Alternative Hypothesis:		One-Sided (H1: AUC > AUC0 or H1: AUC < AUC0)									
Type of Data:		Discrete (Ratings)									
B (SD Ratio):		1									
FPR Limits:		Lower = 0, Upper = 1									
Area Under the ROC Curve											
Power		Sample Size			Actual			Adjusted for FPR Limits			
Target	Actual	N+	N-	N	H0 AUC0	H1 AUC1	Difference Diff	H0 AUC0'	H1 AUC1'	Difference Diff'	Alpha
③ 0.9	0.90085	171	171	342	① 0.8	② 0.88	0.08	0.8	0.88	0.08	④ 0.025
Target Power		The desired power value (or values) entered in the procedure. Power is the probability of rejecting a false null hypothesis.									
Actual Power		The power obtained in this scenario. Because N+ and N- are discrete, this value is often (slightly) larger than the target power.									
N+ and N-		The number of items sampled from each population.									
N		The total sample size. $N = (N+) + (N-)$.									
AUC0 and AUC1		The actual areas under the ROC curve for the null and alternative hypotheses, respectively.									
Diff		The difference to be detected. $\text{Diff} = \text{AUC1} - \text{AUC0}$.									
AUC0' and AUC1'		The adjusted (rescaled to account for FPR limits) areas under the ROC curve for the null and alternative hypotheses, respectively.									
Diff'		The adjusted (rescaled) difference to be detected. $\text{Diff} = \text{AUC1}' - \text{AUC0}'$.									
Alpha		The probability of rejecting a true null hypothesis.									
B (SD Ratio)		The ratio of the standard deviations of the negative and positive groups. $B = \text{SD-} / \text{SD+}$.									
FPR Limits		The lower and upper bounds on the false positive rates.									

그림 5. 대상자수 산출 프로그램의 예

7) 의약품 임상시험 통계가이드라인에 의하면, “제 2종 오류는 일반적으로 10%에서 20%로 정해져 있음”을 기술하였습니다. 검정력의 정의는 (1-제 2종의 오류)로, 자세한 내용은 [부록 3. 주요용어]를 참고하시기 바랍니다.

가정	가정 값	[그림 5] 프로그램 상의 의미
참고 성능효과 (기준 AUROC)	0.80	① H0 AUC0
목표 성능효과 (AUROC)	0.88	② H1 AUC1
검정력	0.9 (90%)	③ Power(Target)
유의수준	0.025 (단측 2.5%)	④ Alpha (Alternative hypothesis 참고)
중도탈락률	0.2 (20%)	

표 3. 대상자수 산출결과에 대한 프로그램 설명

중도탈락을 제외한 유효례수는 총 342명(음성과 양성인 대상자 각각 171명씩)이 필요한 것으로 계산되었습니다. 중도탈락률 0.2를 가정하게 되면, $(171/(1-0.2)) \times 2 = 428$ 명(소수점 올림)으로, 최종 428명을 목표로 임상시험을 설계하게 됩니다.

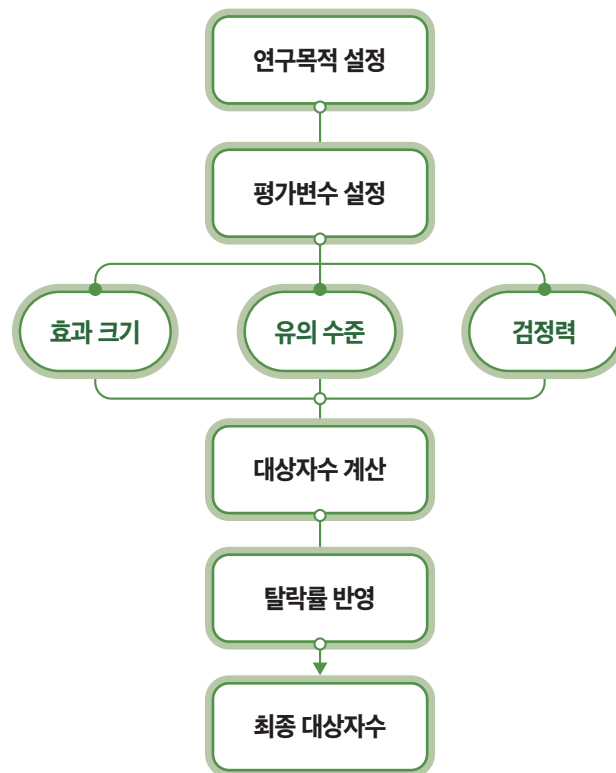


그림 6. 대상자수 산정 흐름도

대상자수 산정을 위해서는 목표 유효성/성능 크기와 참고 유효성/성능 크기, 검정력과 유의수준, 그리고 중도탈락 등이 필요합니다. 유의수준은 유효성/성능 효과가 없을 때, 유효성/성능 효과가 있다고 판단할 오류의 한계치로, 양측 5%(혹은, 단측 2.5%)를 일반적으로 사용합니다. 검정력은 유효성/성능 효과가 있을 때, 유효성/성능 효과가 있다고 판단할 확률로 일반적으로 80% 이상을 가정합니다.

2.3 | 대조군 선정

대조군(Control)은 시험군(Treatment, 임상시험 평가대상에 해당되는 의료제품을 사용하는 군)에서 관측된 결과가 의료제품 자체의 효과에 의한 것인지를 평가하기 위해 기준이 되는 집단을 의미합니다. 적절한 대조군이 설정되지 않은 경우, 관측 결과가 의료제품의 효과인지, 자연적인 변화에 의한 것인지, 또는 (통제되지 않은) 다른 요인에 의한 것인지를 구분하기 어려울 수 있습니다. 따라서 대조군 선정은 임상시험 결과의 신뢰성과 해석 가능성에 직접적인 영향을 미치는 중요한 설계 요소 중 하나입니다.

대조군은 의료제품의 특성이나 사용목적, 기존 치료법 또는 허가제품 여부 등을 종합적으로 고려하여 선정합니다. 기존에 동일목적으로 기존 의료제품이나 표준치료가 사용되고 있는 경우 해당 의료제품 또는 치료법을 대조군으로 설정하는 것이 일반적입니다. 이러한 경우 임상시험은 목표 의료제품(치료군)이 경쟁 의료제품(대조군)과 비교하여 더 우수한지 또는 적어도 동등한 수준의 유효성/성능을 보이는지를 평가하게 됩니다.

반면, 임상시험에서 비교 가능한 의료제품이나 표준치료가 존재하지 않거나 특성상 대조군 설정이 어려운 경우에는 단일군(single-arm)⁸⁾ 임상시험이 고려될 수 있습니다. 앞 절의 예시처럼, 단일군 의료기기 임상시험에서는 사전에 설정된 참고 기준이나 문헌 결과 등을 활용하여 의료기기의 성능을 평가할 수 있습니다. 즉, 선행 임상시험 또는 진료 자료로부터 계획된 임상시험의 선정/제외기준에 적합한 대상자를 후향적으로 선별하는 임상시험을 설계할 수 있습니다. 이후 선별된 자료를 시험군 의료기기에 적용하여 그 결과를 기존 결과와 비교하고, 결과 일치 여부를 평가하여 유효성을 확인하는 임상시험이 가능합니다. 이때, 참고 성능효과(참고치)를 선행된 유사 임상시험들의 AUROC 결과들의 가중평균으로 설정하여 가정할 수 있고, 시험군 의료기기 성능이 참고치 보다 높음(우월성)을 입증하는 것으로 임상시험 목적을 설정할 수 있습니다.

실무적으로는 대조군 선정의 타당성이 임상시험 결과의 규제적인 수용성 측면에 영향을 미칠 수 있으므로 대조군 선정 근거를 명확히 제시하는 것이 중요합니다. 특히 기존 허가제품이 존재함에도 불구하고, 해당 제품과 비교하지 않고 단일군 연구를 수행하고자 한다면 그 필요성과 타당성에 대한 충분한 설명이 요구될 수 있습니다.

이와같이 대조군은 시험 의료제품의 임상적 유효성/성능효과를 객관적으로 평가하기 위한 기준점 역할을 하므로 의료제품의 특성과 임상시험 목적에 적합한 대조군을 선정하는 것이 중요합니다. 또한, 대조군 선정은 평가변수, 통계분석방법 및 대상자수 산정에 영향을 미칠 수 있으므로 임상시험 설계 초기 단계에서 임상전문가와 통계전문가 및 규제 담당자와 충분히 논의하여 결정하는 것이 바람직합니다.

8) 연구 목적에 따라 동일 시험대상자에 대해 중재 전후를 비교하거나, 동일 검체에 대해 시험기기와 참조시험(또는 대조기기)로 모두 검사하여 성능을 비교하는 대응쌍 설계(paired design)도 단일군 시험에 포함됩니다. 대응쌍 설계에 대한 자세한 설명은 안내서 범위에서 생략하였습니다.

시험군은 임상시험에서 평가하고자 하는 의료제품을 적용한 군이고, 대조군은 시험군의 결과와 비교 평가하기 위해 기준이 되는 군입니다. 대조군을 포함하는 비교연구에서 대조군 선정 근거에 대해, 또는 단일군 연구인 경우 단일군 설정 타당성에 대한 설명이 필요할 수 있습니다.

2.4 | 무작위배정과 눈가림

무작위배정(Randomization)과 눈가림(Blinding)은 임상시험 결과의 신뢰성을 높이고 편향(Bias)을 최소화 하기 위해 사용되는 대표적인 임상시험 설계 요소입니다. 이 두가지 요소는 실제 효과를 보다 객관적으로 평가하기 위한 목적으로 사용됩니다.

무작위배정은 대상자를 시험군 또는 대조군에 우연에 의해 배정하는 방법입니다. 이를 통해 연구자나 시험 대상자의 의도에 따라 특정 군에 대상자가 집중되는 것을 방지할 수 있습니다. 예를 들어, 임상시험에 참여한 시험대상자의 질환 중등도가 시험군과 대조군 간에 균형이 깨진다면, 결과가 왜곡될 수 있습니다. 무작위배정은 이러한 차이를 최소화하기 위해 시험군과 대조군이 가능한 유사특성을 갖도록 하는 데 목적이 있습니다. 무작위배정이 적절히 수행되면 두 군 간 결과의 차이를 의료제품 자체의 효과로 해석할 수 있는 근거가 강화됩니다. 실무적으로는 단순 무작위배정, 블록 무작위배정, 층화 무작위배정 등의 방법이 있으며, 대상자수가 적거나 다기관 연구인 경우에도 군 간 균형을 유지하기 위해 적절히 사용될 수 있습니다. 이는 대상자수, 질환특성 및 연구목적 등을 고려하여 결정합니다.

눈가림은 임상시험에 참여하는 시험대상자, 연구자 또는 결과평가자가 군배정 정보를 알수 없도록 하는 방법입니다. 눈가림의 목적은 기대효과나 주관적 판단 때문에 결과에 영향을 끼치는 것을 최소화 하는 데 있습니다. 예를 들어, 시험대상자가 자신이 시험군의 의료제품을 투여받거나 혹은 사용하고 있다는 사실을 알고 있는 경우, 실제 유효성/성능효과와 무관하게 증상이 개선되었다고 응답할 수 있습니다. 또한 연구자가 군배정 정보를 알고 있는 경우 결과 평가에 무의식적인 영향을 받을 가능성이 있습니다. 이러한 영향을 줄이기 위해 단일눈가림(single-blind), 이중눈가림(double-blind) 등의 방법이 사용됩니다.

무작위배정과 눈가림은 임상시험의 내부타당성을 높이는 중요한 방법이지만, 모든 의료제품 임상시험에 적용되는 것은 아닙니다. 예를 들어, 수술용/시술형 의료기기나 물리적 특성이 명확히 구분되는 의료기기의 경우 시험대상자나 연구자에 대한 눈가림이 현실적으로 어려울 수 있습니다. 이런 경우 독립적 평가자를 활용하거나 객관적 평가변수를 설정하는 등의 편향을 최소화하기 위한 대안을 고려할 수 있습니다. 무작위배정방법, 배정절차 및 눈가림 적용 여부는 임상시험 설계단계에서 사전에 정의되어야 하며, 실제 수행 과정도 계획된 절차에 따라 적절하게 유지되도록 관리하여야 합니다.

무작위배정과 눈가림은 임상시험 결과의 신뢰성을 높이고 편향을 최소화하기 위한 중요한 설계방법으로서 임상시험에서 의료제품의 특성과 연구목적에 고려하여 적절하게 적용 여부를 검토할 필요가 있습니다.

무작위배정은 시험군-대조군 비교 임상시험에서 시험대상자를 각 군에 무작위로 배정하는 방법으로, 체외진단의료기기 가이드라인⁹⁾에서 설명하는 단일군 임상시험의 무작위 표집(random sampling, 임상시험대상자나 검체를 모집단에서 무작위로 선정)과는 구분됩니다.

동일한 의료제품을 평가하더라도 증상이 심한 환자가 시험군에 더 많이 배정되면 결과가 실제보다 나쁘게 나타날 수 있습니다. 반대로 증상이 가벼운 환자가 시험군에 집중되면 결과가 실제보다 좋게 나타날 수 있습니다. 따라서, 기저시점에서 임상시험대상자들 특성 상태가 시험군과 대조군 간 통계적으로 차이가 없어야 하고, 이를 위한 설계 요소로 무작위배정이 사용됩니다.

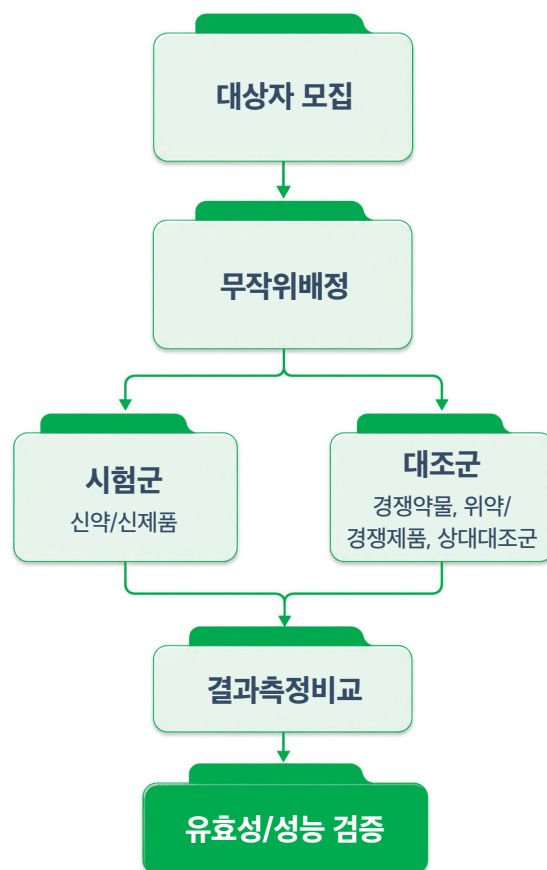


그림 7. 시험군과 대조군의 구성 및 검증과정

9) 체외진단의료기기 임상적 성능시험 시 통계적 고려사항(민원인 안내서), 2023

2.5 | 결측치 및 중도탈락

결측치(Missing data)와 중도탈락(Drop-out)은 모든 임상시험에서 발생할 수 있는 전형적 현상입니다. 결측치는 필요한 자료가 계획된 평가시점에 수집되지 않은 상태를 의미하고, 중도탈락은 대상자가 임상시험 종료 이전에 임상시험 참여를 중단하여 이후 자료 수집이 불가능한 상태를 의미합니다. 두 현상은 서로 연관성이 있으나 동일한 개념은 아닙니다. 결측치는 자료 수준의 문제이고, 중도탈락은 시험대상자 수준의 문제입니다.

의료제품 임상시험에서는 시험대상자의 동의철회, 이상반응, 혹은 추적 실패 등으로 중도탈락이 발생할 수 있습니다. 또한, 의료기기 임상시험에서도 기기 사용과정에서 불편성이나 순응도 저하, 기대효과 미충족 또는 이상반응 발생등의 요인으로 중도탈락이 발생할 수 있습니다. 반복측정이 필요한 임상시험 설계인 경우, 특정 평가시점에서 측정 누락으로 결측이 발생할 수 있습니다.

결측치나 중도탈락이 적절하게 관리되지 않을 경우, 최종분석대상 수가 감소하게 되어 통계 검정력 저하를 초래할 수 있고, 특정 군에서 탈락이 집중되는 경우에는 결과의 편향이 발생할 수 있습니다. 특히, 효과가 낮은 시험대상자의 중도탈락이 많은 경우 의료제품의 유효성/성능효과가 실제보다 과대평가될 위험이 있습니다. 따라서, 임상시험 설계단계에서는 예상 중도탈락률을 고려하여 대상자수를 산정하고, 수행단계에서는 결측발생이 최소화되기 위하여 자료수집절차와 자료관리계획을 수립하는 것이 중요합니다.

결측치나 중도탈락 처리방법은 임상시험 계획서에 사전에 명확히 정의되어야 합니다. 일반적으로 모든 시험대상자를 포함하여 분석하는 ITT(Intention -To-Treat) 원칙에 기반한 FAS(Full Analysis Set) 분석과 실제 프로토콜을 충실히 준수한 시험대상자만을 분석하는 PPS(Per-Protocol Set) 분석이 함께 활용되며, 두 분석 결과를 종합적으로 검토하는 것을 권장합니다.

「체외진단의료기기 임상적 성능시험 시 통계적 고려사항(민원인 안내서)」에 의하면, 임상적 성능평가 시험의 경우에는 검체 수집이 불가능하거나, 검체의 변질 또는 오염, 시험기기의 이상 등으로 인하여 시험기기 검사결과의 결측이 발생할 수 있습니다. 이러한 경우에도 전체 검체 중 측정 (불)가능한 검체 수와 그 사유(검체변질, 오염, 검사법 오류, 검체누락 등)을 정리요약하고, 측정 불가능한 자료의 처리 방법에 대한 방법을 사전에 계획하여야 합니다.

결측 및 중도탈락 발생가능성을 고려하여, 예상 중도탈락률은 대상자수 산정에 반영되어야 하고, 이를 최소화하기 위한 자료관리방안 및 추적관찰 계획, 그리고 결과의 완건성을 위한 결측처리와 분석원칙을 사전에 정의하여야 합니다.

2.6절과 2.7절에서는 각각 의약품 임상시험과 의료기기 임상시험 예시를 통해 통계적 핵심 요소의 설계가 어떻게 기술될 수 있는지를 살펴보고자 합니다.

2.6 | [의약품 예시] 본태성 고혈압 환자 대상의 유효성 평가

임상시험 목적 및 설계

본태성 고혈압 환자를 대상으로 복합제 ☆ 의 유효성과 안전성 평가를 위한 다기관, 무작위배정, 이중눈가림, 활성대조, 우월성, 제 3상 임상시험

▶ 평가변수

- 1차 유효성 평가변수: 기저대비 의약품 투여 8주 시점의 MSSBP 변화량
- 2차 유효성 평가변수:
 - 기저 대비 의약품 투여 4주 시점의 MSSBP 변화량
 - 기저 대비 의약품 투여 4주, 8주 시점의 MSDBP 변화량
 - 기저 대비 의약품 투여 4주, 8주 시점의 맥압 변화량
 - 기저 대비 의약품 투여 4주, 8주 시점의 혈압정상화 비율
- 안전성 평가변수:
 - 이상사례 (TEAE)
 - 활력징후
 - 실험실적 검사
 - 심전도 검사
 - 신체검진

▶ 대상자 수: 총 436명 (군당 218명)

- 설정근거¹⁰⁾
 - 복합제 ☆ 의 기저대비 8주 MSSBP 변화량: -16.5
 - 복합제 ☆ 의 기저대비 8주 MSSBP 변화량에 대한 표준편차: 12.09

10) 대상자수 산출을 위한 통계적 가정값은 학습을 위한 예시이며, 연구목적 및 참고근거에 따라 적절하게 설정해야 합니다.

- 단일제 © 의 기저대비 8주 MSSBP 변화량: -12.9
- 단일제 © 의 기저대비 8주 MSSBP 변화량에 대한 표준편차: 11.8
- (양측) 유의수준: 0.05
- 검정력: 0.8

- 목표 표본수: 348명 (군당 174명)
- 중도탈락률 20%(0.2) 고려된 최종 표본수
 - 총 436명 (군당 218명)

Numeric Results

Solve For: **Sample Size**
 Test Type: Two-Sample Welch's Unequal-Variance T-Test
 Difference: $\delta = \mu_1 - \mu_2$
 Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$

Power		Sample Size			Mean			Standard Deviation		Alpha
Target	Actual	N1	N2	N	μ_1	μ_2	Difference δ	σ_1	σ_2	
0.8	0.80042	174	174	348	-16.5	-12.9	-3.6	12.09	11.8	0.05

Target Power The desired power value (or values) entered in the procedure. Power is the probability of rejecting a false null hypothesis.
 Actual Power The power obtained in this scenario. Because N1 and N2 are discrete, this value is often (slightly) larger than the target power.
 N1 and N2 The number of items sampled from each population.
 N The total sample size. $N = N_1 + N_2$.
 μ_1 and μ_2 The assumed population means.
 δ The difference between population means at which power and sample size calculations are made. $\delta = \mu_1 - \mu_2$.
 σ_1 and σ_2 The assumed population standard deviations for groups 1 and 2, respectively.
 Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A parallel two-group design will be used to test whether the Group 1 mean is different from the Group 2 mean ($H_0: \mu_1 - \mu_2 = 0$ versus $H_1: \mu_1 - \mu_2 \neq 0$). The comparison will be made using a two-sided, two-sample, unequal-variance t-test (Welch, 1938), with a Type I error rate (α) of 0.05. The standard deviation for Group 1 is assumed to be 12.09 and the standard deviation for Group 2 is assumed to be 11.8. To detect a difference in means of $\mu_1 - \mu_2 = -16.5 - -12.9 = -3.6$ with 80% power, the number of needed subjects will be 174 in Group 1 and 174 in Group 2. The sample size was computed using PASS 2025, version 25.0.3.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
	N1	N2	N	N1'	N2'	N'	D1	D2	D
20%	174	174	348	218	218	436	44	44	88

▶ 대조군

단일제 ©

◆ 무작위배정

본 임상시험은 과학적 타당성을 보장하고, 치료군 배정에 시험자의 주관이 개입되지 않도록 하기 위하여 이중눈가림, 무작위배정으로 진행된다. 본 임상시험에서 시험대상자는 시험군 또는 대조군에 1:1의 비율로 무작위배정되며, 이때 무작위배정은 베이스라인 MSSBP(160 mmHg 미만 또는 이상)를 층화요인으로 하는 층화 무작위배정 방법을 이용한다.

본 임상시험 참여에 서면 동의한 시험대상자에게 시험기관 내 동의받은 순서대로 스크리닝 번호를 부여한다. ■ 주간의 위약도입기 및 생활습관교정을 수행 후, 무작위배정 시점에 최종적으로 본 임상시험 참여에 적합하다고 판단된 시험대상자는 등록을 결정한 순서대로 Interactive Web Response System(IWRS)를 통해 무작위배정 번호와 임상시험용 의약품 번호를 부여받는다.

임상시험용 의약품 번호는 해당 시험대상자에게 불출하는 임상시험용 의약품에 대한 고유 식별번호를 의미한다. 시험대상자는 임상시험용 의약품 번호를 확인하여 배정된 대로 치료기용 임상시험용 의약품을 처방한다.

(중략)

본 임상시험 종료 시까지 임상시험을 수행하는 사람은 모두 눈가림 상태가 유지되어야 하며, 본 임상시험 중 눈가림 해제가 필요한 경우 ‘눈가림 해제’ 절차에 언급된 절차에 따라 진행한다.

◆ 눈가림

본 임상시험은 다기관, 무작위배정, 이중눈가림, 활성대조, 제3상 임상시험으로, 시험대상자와 시험자의 주관적인 판단을 배제하기 위하여 이중눈가림으로 진행되며, 시험자 혹은 시험대상자의 치료 효과 및 이상사례 등을 평가하는 데 비뚤림을 방지하기 위하여 임상시험 기간 동안 눈가림을 유지해야 한다. 임상시험 도중 응급상황 발생으로 인하여 해당 대상자의 임상시험용 의약품 확인이 반드시 요구되는 경우, 시험자는 즉시 스폰서 또는 스폰서가 위임한 임상시험 수탁기관의 임상시험모니터요원(이하 ‘모니터요원’)에게 해당 사실을 알려야 한다. 이때, 눈가림 해제로 확인되는 무작위배정 코드는 가능한 최소한으로만 공개되어야 한다. 눈가림이 해제된 경우 사유와 해제일에 대해 눈가림 해제 문서를 기록 및 보관해야 한다. 무작위배정 코드가 해제된 대상자는 임상시험을 지속할 수 없다. 임상시험 종료 후 모든 데이터에 대하여 쿼리를 통해 문제점을 해결한다. 데이터베이스가 완전하고, 정확함을 확인한 후 데이터베이스 잠금을 실시하고, 치료기용 임상시험용 의약품의 무작위배정 코드가 공개될 것이다. 그 이후의 데이터베이스의 변경은 의뢰자와 데이터베이스 관리자의 서면 동의하에서만 가능하다.

▶ 결측치

유효성 평가변수 분석 시, 중도탈락 등의 사유로 인하여 결측치가 발생하여 결측치 처리가 필요할 경우에는 결측치 발생 이전 시점의 가장 마지막에 얻은 자료를 해당시점에서 얻은 자료로 대체하는 LOCF(Last Observation Carried Forward)¹¹⁾ 방법을 적용한다. 단, 결측치 처리는 FAS 군에만 적용하며, 안전성 평가변수에 대해 결측 자료가 발생한 경우 보정 없이 관찰된 값을 그대로 이용하여 분석한다.

그 외 임상시험 계획서 본문에 포함되는 통계 사항의 예¹²⁾는 다음과 같습니다.

▶ 통계 분석방법

● 1차 유효성 평가방법

투여군별 기저치 및 8주 시점에서의 MSSBP 측정값과 기저 대비 8주 후 변화량에 대하여 기술통계량(대상자 수, 평균, 표준편차, 중앙값, 최솟값, 최댓값)을 제시하고, 시험군과 대조군간 변화량의 차이는 기저치를 공변량으로 하는 공분산분석(ANCOVA)로 분석한다.

● 2차 유효성 평가방법

MSSBP 변화량의 경우 1차 유효성 평가와 동일한 방법을 사용하여 분석한다. MSDBP와 맥압의 경우, 기저치(Baseline) 및 각 시점에서의 측정치에 대해 투여군별 기술통계량(시험대상자 수, 평균, 표준편차, 중앙값, 최솟값, 최댓값)을 제시한다. 시험군과 대조군간의 기저대비 임상시험용 의약품 투여 후 각 시점에서의 변화량에 대하여 각 평가변수의 기저치와 총화요인을 공변량으로 하는 공분산분석(ANCOVA)을 실시한다. ANCOVA 결과는 각 투여군별 최소제곱평균(LSM)과 표준오차(SE), 투여군간 LSM 차이에 대응되는 양측 95% 신뢰구간, p-value로 요약한다. 혈압 정상화 비율에 대해서는 각 시점에서의 빈도와 백분율을 투여군별로 제시하고, 시험군과 대조군간 차이에 대해서는 총화요인의 효과를 보정한 Cochran-Mantel-Haenszel test (CMH test)로 분석한다.

● 안전성 분석방법 (신체검진의 예)

스크리닝(Visit 1) 및 무작위배정(Visit 2) 시점에 신체검진 결과를 목록화하고, 각 투여군별로 기저시점에서는 정상(Normal) 또는 임상적으로 의미 없는 비정상(Abnormal NCS)이었다가 투여 후 임상적으로 의미 있는 비정상(Abnormal CS)으로 변화하였는지의 여부에 대한 table을 제시한다. 임상시험용 의약품 투여 이후 각 방문 또는 중도탈락 방문 시 새롭게 나타난 신체검진상 비정상 결과를 목록화하여 제시한다.

11) 결측처리 방법에는 LOCF 외 여러 방법이 있으나 본 안내서의 범위를 벗어나므로 자세한 설명은 생략하였습니다.

12) 발체하는 과정에서 안내서 내용에서 벗어나지 않는 범위 내에서 수정 또는 축약될 수 있는 예시 사례입니다.

▶ 분석군

모든 유효성 분석은 Full Analysis Set(이하 'FAS')에 대한 분석을 주 분석으로 하고, Per-Protocol Set(이하 'PPS')에 대한 분석을 보조적으로 시행한다.

● Full Analysis Set (FAS)

FAS는 무작위배정 이후 임상시험용 의약품을 한 번 이상 투여받았고, 베이스라인 및 베이스라인 후 일차 유효성 평가 변수인 MSSBP가 한 번 이상 측정된 대상자를 포함한다. FAS는 무작위배정된 투여군에 따라 분석한다.

● Per-Protocol Set(PPS)

PPS는 FAS에 포함되는 대상자 중 중대한 계획서 위반 없이 임상시험을 완료한 대상자를 포함한다. PPS에서 제외되는 중대한 계획서 위반은 다음과 같으며, 그 외 중대한 위반 사항은 무작위배정 코드 공개 전에 (Blind) 분석군 회의 등을 통해 결정한다.

(중략)

● Safety set (SS)

임상시험용 의약품을 한 번 이상 투여받은 모든 대상자를 포함한다. SS는 실제 투여군에 따라 분석한다.

▶ 중간분석

중간분석은 계획되지 않았으며, 임상시험이 종료된 후 분석을 실시할 것이다.

2.7 | [의료기기 예시] CT영상 기반의 질환 위험도 분석에 대한 유효성 평가

디지털의료기기 임상시험등의 경우에는 참조기준(Gold standard) 또는 유사 기기의 선행 임상시험 결과와 비교하여 기준을 설정하기도 합니다. 이 경우, 자료 수집이나 처리 방법에 따라 전향적 혹은 후향적 시험설계가 가능합니다. 전향적 임상시험은 임상적 증거의 수준에서 매우 긍정적인 측면이 있으나 많은 시간과 비용이 소모되는 단점이 있어 특히 희귀 질환 또는 임상적 상태를 확인하는 데 오랜 시간이 필요한 질환에서는 효율성이 떨어집니다. 반면에 후향적 임상시험은 시간과 비용을 절약하는 장점을 가지고 있으나, 다양한 편향이 발생할 가능성이 상대적으로 높은 단점을 가지고 있어 연구 계획과 수행에 주의하여야 합니다.

앞서 설명된 통계 개념들을 의료기기 임상시험의 예시를 통해 한번 더 검토해보는 것도 임상시험 개발에 도움이 될 것입니다. 다음은 이해를 돕기 위한 예시이므로, 실제 임상시험에 그대로 적용하기보다 개발제품의 사용목적 및 임상적 유효성을 고려하여 적절한 임상시험 설계를 마련해야 합니다.

임상시험 목적 및 설계

☆ 진료과 의사가 ☆ 종양이 의심되는 환자의 위험도를 분석하기 위해 사용하는 소프트웨어로
◎ CT영상을 인공지능 모델로 분석하여 종양의 양성/악성 위험도 분석에 대한 임상적 유효성을 평가하기 위한 후향적, 단일기관, 개방, 확증적 임상시험

▶ 평가변수

- 1차 유효성 평가변수: 민감도(Sensitivity), 특이도(Specificity)
- 2차 유효성 평가변수: 양성예측도(Positive Predictive Value, PPV), 음성예측도(Negative Predictive Value, NPV)
- 안전성 평가변수: 해당사항 없음

▶ 표본 수

총 709개 (악성 554개, 양성 155개)

- 민감도 설정근거¹³⁾
 - 시험기기 예측치 0.94
 - Reference value 0.90
 - (양측) 유의수준 0.05
 - 검정력 0.9
- 특이도 설정근거¹⁴⁾
 - 시험기기 예측치 0.86
 - Reference value 0.75
 - (양측) 유의수준 0.05
 - 검정력 0.9
- 목표 표본수
 - 악성 498, 양성 139
- 중도탈락률 10%(0.1) 고려된 최종 표본수
 - 악성 554, 양성 155

13), 14) 학습을 위한 예시값으로 연구목적 및 참고근거에 따라 적절하게 설정하여야 합니다.

Numeric Results

Solve For: [Sample Size](#)
 Test Type: Exact Test
 Alternative Hypothesis: One-Sided ($H_0: P \leq P_0$ vs. $H_1: P > P_0$)

Power*	Sample Size n	Proportion		Alternative Difference P1 - P0	Alpha	Reject H0 If R $\geq$
		Null P0	Alternative P1			
0.90022	498	0.9	0.94	0.04	0.025	462

Numeric Results

Solve For: [Sample Size](#)
 Test Type: Exact Test
 Alternative Hypothesis: One-Sided ($H_0: P \leq P_0$ vs. $H_1: P > P_0$)

Power*	Sample Size n	Proportion		Alternative Difference P1 - P0	Alpha	Reject H0 If R $\geq$
		Null P0	Alternative P1			
0.90176	139	0.75	0.86	0.11	0.025	115

* Power was computed using the normal approximation method.

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
 n The size of the sample drawn from the population.
 P0 The value of the population proportion under the null hypothesis.
 P1 The value of the population proportion under the alternative hypothesis.
 P1 - P0 The difference to be detected by the study.
 Alpha The probability of rejecting a true null hypothesis.
 Reject H0 If... Gives the critical value(s) for the test.

대조군

대조기기 없음

무작위화

본 후향적 임상시험에서는 수집된 전자의료데이터(EMR)를 바탕으로 평가를 수행한다. 최종 선정/제외 기준을 확인한 결과, 본 후향적 임상시험에 적합한 것으로 최종 판단된 데이터를 대상으로 무작위 표집 (random sampling)을 실시한다. 한 명의 대상자로부터 여러 개의 데이터를 추출할 수 있으며, 이 때 추출하는 데이터의 개수는 종양 크기와 촬영된 phase 개수에 따라 달라질 수 있다. 단일 대상자로부터 확인된 CT 이미지 중 최초 이미지와 가장 마지막 이미지를 제외하고, 이미지 간 간격이 ■ 이상이 되는 시험 데이터를 추출하여 데이터 풀을 만든 후 등록 순으로 번호를 부여하고 데이터 서버에 저장한다.

▶ 눈가림

본 임상시험은 편향을 최소화하기 위하여 참조표준 설정에 참여하지 않은 의료기기 관리자가 디지털 의료기기를 적용하여 결과 값을 기록한다. 의료기기 관리자는 참조표준 결과에 대한 눈가림을 적용하도록 설계하였다.

▶ 결측치 처리

유효성 평가 시 참조표준의 결과가 결측인 경우 결측값은 대체하지 않고 있는 자료(available data set)를 그대로 사용하여 분석하며, 시험기기의 결과가 결측인 경우 결측값은 Worst case(실패) 대체하여 분석한다.

▶ 통계 분석방법

● 1차 유효성 평가방법

민감도 및 특이도 결과를 제시하고, 95% 신뢰구간을 제시한다. 이때, 일차 유효성 평가 분석 결과 민감도 및 특이도 각각의 신뢰구간 하한이 0.90, 0.75를 초과하는 경우¹⁵⁾ 시험군의 통계적 유의성을 입증한 것으로 판단한다.

● 2차 유효성 평가방법

양성예측도 및 음성예측도 결과를 제시하고, 95% 신뢰구간을 제시한다.

● 안전성 분석방법

해당사항 없음

▶ 분석대상군

참조표준의 결과를 얻을 수 있는 모든 데이터를 유효성 평가에 포함한다.

15) 학습을 위한 예시값으로 연구목적 및 참고근거에 따라 적절하게 설정하여야 합니다.



본 장에서는 임상통계 전문인력이 제한적인 경우, 임상시험을 기획하거나 수행과정에서 외부전문기관과의 협업이나 설계 초기단계, 그리고 규제기관 지원 방안에 대해서 안내하고자 합니다.

3.1 | CRO 활용 시 확인 사항

임상시험 수행과정에서 CRO를 활용하는 경우, 업무위탁을 위해서는 의료제품 임상시험의 경험여부나, 규제대응 경험, 통계전문인력 포함여부 등을 파악하겠지만, 이뿐만 아니라 임상시험 설계 및 분석 전반에 대한 이해와 관리도 필요합니다. 즉, 다음과 같은 주요 사항을 통계전문가 또는 CRO와 함께 검토하되, 시험의뢰자는 그 의미를 이해하고 주요 가정과 설계의 타당성을 검토할 필요가 있습니다.

- 임상시험 목적과 계획서 내 통계분석방법이 일관되었는지
- 평가변수와 분석결과가 임상적으로 이해가능한지
- 대상자수 산정 근거 설명이 적절하게 제시되었는지
- 결측치와 중도탈락 처리방법이 사전에 명확히 정의되어 있는지
- 분석대상군 정의가 명확히 설명되었는지
- (분석 후) 분석결과가 의사결정에 활용가능한 형태로 제시되었는지¹⁶⁾

3.2 | 임상시험 설계 단계 권고사항

임상시험의 성공적인 수행을 위해서 임상시험 설계 초기 단계부터 임상적 평가전략을 고려하는 것이 중요합니다. 설계 단계에서의 의사결정은 평가변수 설정, 대상자수 산정, 대조군 선정 및 통계분석방법에 직접적인 영향을 미칠 수 있습니다. 임상시험 설계 단계에서는 다음과 같은 사항을 검토하는 것을 권장합니다.

16) 주요 용어 중, 유의확률(p-value)와 신뢰구간(CI, confidence interval)을 참고하기 바랍니다.

항목	확인사항	
임상시험 목적	임상시험 목적이 명확하게 기술되었는가?	☐
임상시험 설계	임상시험 계획서에 임상시험이 단일군인지, 비교시험인지 등 설계에 대한 정의가 명확히 기술되었는가? (예. 후향적, 단일기관, 확증임상시험 등)	☐
대조군	비교제품이나 (의료기기의 경우) 참조값(Gold Standard)이 명확하고 근거제시가 되어 있는가?	☐
일차평가변수	1차 평가변수가 명확하고 측정방법이 정의 되어 있는가?	☐
대상자수/ 표본수 산출	유의수준, 검정력, 효과크기 기반으로 표본 수가 산출되었는가? 이때, 효과크기 가정 시 근거제시는 되었는가?	☐
편향 통제	무작위배정, 눈가림 방법이 포함되었는가?	☐
통계분석 방법	자료 분석에 대한 일반적 원칙이 있는가? 1차 평가변수에 대한 분석방법이 명확히 기술되었는가? 그 외 2차 평가변수 또는 탐색/안전성 평가변수에 대한 분석방법 기술이 있는가?	☐
결측치 처리	결측치 처리 방법에 대해 기술되었는가? (결측처리 없이 분석하고자 할 때에도 그에 대한 근거가 제시되었는가?)	☐

표 6. 임상시험 계획서 작성 시 통계 관련 체크리스트

이와 같은 사전 검토가 충분히 이루어지지 않는 상태에서 임상시험을 설계할 경우, 일차평가변수 변경 등의 추가자료 확보가 필요하게 되어 일정이 지연될 수 있습니다. 따라서 임상시험 설계는 임상적, 규제적 요구사항을 함께 고려하여 통합적으로 설계하는 것이 바람직합니다.

3.3 | 규제기관 활용

의료제품 임상시험 과정에서 규제기관과의 적절한 사전 소통은 불확실성을 줄이고 임상시험의 방향성을 명확히 하는데 중요한 역할을 합니다. 사전상담 제도를 통해 임상시험 설계, 평가변수의 설정, 대조군 선정, 통계분석 방법 등에 대한 의견을 제공할 수 있습니다. 다음과 같은 상황에서 규제기관 지원을 활용하는 것이 특히 유용할 수 있습니다.

- 새로운 유형의 의료제품 또는 기존 사례가 부족한 경우
- 평가변수 또는 성능기준 설정이 어려운 경우
- 단일군 설계 또는 외부 대조군 활용이 필요한 경우
- AI 기반 또는 소프트웨어 의료기기 등 새로운 분석 방법이 적용되는 경우

사전상담을 통해 임상시험 설계 단계에서 규제 기대 수준을 확인하면 이후 허가 심사과정에서 발생할 수 있는 불확실성을 줄이고 개발 효율성을 높일 수 있습니다. 결과적으로 규제기관과의 사전 소통은 선택이 아닌, 임상시험 성공 가능성을 높이기 위한 중요한 수단으로 활용될 수 있습니다.

부록 1. 주요 통계 가이드라인

임상시험의 설계 시 다음과 같은 통계관련 가이드라인을 참고할 수 있습니다.

- 의약품 임상시험 통계 가이드라인, 식품의약품안전처 식품의약품안전평가원, 2016
- 의료기기 임상시험 관련 통계기법 가이드라인. 식품의약품안전처 식품의약품안전평가원, 2010
- 체외진단의료기기 임상적 성능시험 시 통계적 고려사항(민원인 안내서), 2023

부록 2. 평가변수별 주요 통계 분석법

평가변수 유형	평가변수 예시	주요 통계 분석법
연속형	■ 기저대비 평가시점 척도/점수 변화량 ■ 정량제품 일치도 등	■ 비열등성 추정값 및 신뢰구간 ■ Independent t-test, Wilcoxon rank sum test ■ ANCOVA, MMRM ■ 대응쌍 비교(Paired t-test, Wilcoxon signed rank test) ■ Passing-Bablok 또는 Deming 회귀분석 및 Bland-Altman 분석
범주형	■ 반응률, 치료성공률 ■ 임상적 민감도/특이도, 양성/음성/전체 일치율 등	■ 비열등성 추정값 및 신뢰구간 ■ Pearson's chi-square test ■ Fisher's exact test ■ Cochran-Mantel-Haenszel test ■ 단일군 이항검정, 대응쌍 비교(McNemar's test)

평가변수 유형	평가변수 예시	주요 통계 분석법
생존형	■ 전체생존기간(OS) ■ 무진행생존기간(PFS)	■ Kaplan-Meier estimation ■ Log-rank test ■ Cox's PH model
연속형 (PK)	AUC, Cmax 등	■ 동등성 추정값 및 신뢰구간

부록 3. 주요 용어

본 절에서는 의료기기 임상시험을 위주로 설명하였으며, 주요용어는 알파벳순·가나다순이 아닌, 개념상 연관성 높은 용어끼리 함께 안내하였습니다. 원하는 용어를 찾을 때는 [찾아보기]를 활용하여 확인할 수 있습니다.

■ 민감도, Sensitivity

민감도는 질병있는 대상자를 검사했을 때, 검사기기가 실제 환자를 얼마나 잘 찾아내는지를 나타내는 값으로, 민감도가 높을수록 질병있는 대상자를 놓칠 가능성이 낮다고 해석합니다. 즉, 실제 환자 중 검사결과를 양성으로 정확하게 판정할 비율로, 공식은 다음과 같습니다.

- 민감도 (Sensitivity) = (True Positive)/(True Positive + False Negative)

민감도가 높을수록 위음성(false negative)이 낮습니다. 예를 들어 100명의 대상자 중 실제 질환자가 80이고, 의료기기가 이 중 72명을 양성(positive)으로 판정했다면 민감도는 $72/80 = 90\%$ 이며 실제 환자 10명 중 9명을 정확히 찾아낸다는 의미입니다.

■ 특이도, Specificity

특이도는 질병없는 대상자를 검사했을 때, 검사기기가 건강한 사람을 정상으로 얼마나 정확하게 판정하는지를 나타내는 값으로, 특이도가 높을수록 정상인을 환자로 잘못판단할 가능성이 낮습니다. 즉, 정상인을 환자로 잘못판단하지 않는 정도를 나타내는 지표로, 공식은 다음과 같습니다.

- 특이도 (Specificity) = (True Negative)/(True Negative + False Positive)

특이도가 높을수록 건강한 사람을 환자로 잘못 판단하는 경우가 낮습니다. 예를 들어, 특이도가 95%라면, 실제로 질병없는 사람 100명 중 95명을 정상으로 올바르게 판정한다는 의미입니다.

민감도가 높을수록 환자를 놓치는 경우가 줄어들고 특이도가 높을수록 정상인을 환자로 잘못 판단하는 경우가 줄어듭니다. 두 지표는 의료기기 진단 성능을 평가할 때 함께 사용되며, 검사 결과의 정확성을 이해하는데 중요한 기준이 됩니다.

■ AUROC, Area Under the Receiver Operating Characteristic Curve

AUROC는 의료기기가 질환이 있는 대상자와 없는 대상자를 얼마나 잘 구분하는지를 평가하기 위한 성능 지표입니다. ROC 곡선을 이용한 평가 방법은 cut-off 값에 대한 민감도와 특이도 간의 연관성을 나타내는 곡선이며, AUROC는 해당곡선 아래 면적으로 단일 측정치(예. 민감도, 특이도 등)를 이용한 평가 방법보다 더 많은 정보를 제공합니다. AUROC 값은 0에서 1 사이의 값을 가지며, 일반적으로 1에 가까울수록 판별 성능이 우수하다고 해석할 수 있습니다.

■ 단일군 연구, Single arm study

비교 대상이 없이 단일 시험군의 유효성/성능효과를 평가하는 시험입니다.

■ 비교연구, Comparative study

임상시험에서 시험군과 대조군을 비교하여 시험군에 해당하는 의료제품을 평가하는 연구입니다.

■ 비열등성 시험, Non-inferiority trial

비교연구에 있어 대조군에 비하여 시험군의 유효성/성능효과가 열등하지 않음을 보이는 연구시험입니다.

■ 우월성 시험, Superiority trial

비교연구에 있어 대조군에 비하여 시험군의 유효성/성능효과가 우월함을 확인하기 위한 시험입니다.

■ 검정력, Power

검정력은 실제로 차이가 존재할 때 통계분석을 통해 그 차이를 올바르게 발견할 수 있는 확률을 의미합니다. 새로운 의료제품이 기존 의료제품보다 성능이 향상되었음에도 임상시험에 참여한 대상자수가 너무 적거나 자료의 변동이 큰 경우에는 실제 차이가 있어도 통계적으로 확인을 못할 수가 있는데, 이런 상황에서 실제 차이를 얼마나 잘 찾아낼 수 있는지를 나타내는 지표가 검정력입니다. 실제 유효성/성능효과가 존재할 때 이를 발견할 가능성이 높을수록 검정력이 높다고 하며, 임상시험 설계 시 필요한 개념으로, 일반적으로 80%(0.8) 이상으로 설정합니다. 즉, 실제 차이가 존재할 경우, 80% 이상의 확률로 그 차이를 발견할 수 있도록 임상시험을 계획한다는 의미입니다.

■ 유의수준, Confidence level

유의수준은 통계분석 결과를 얼마나 엄격하게 판단할 것인지를 정하는 기준입니다. 새로운 의료제품이 기존 의료제품보다 성능이 더 좋다고 분석결과가 나왔다고 가정합니다. 하지만 실제 차이가 없는데도 우연히 차이가 있는 것처럼 보일 수 있습니다. 이러한 “우연에 의한 판단오류”를 어느정도까지 허용할지를 나타낸 것이 유의수준으로 일반적으로 5%(0.05)를 사용합니다. 이는 실제로는 차이가 없는데도 분석 결과를 잘못해서 “차이가 있다”고 결론 내릴 가능성을 5% 미만으로 제한하겠다는 의미입니다. 통계 분석결과로 계산되는 유의확률(p-value)이 판단 기준인 유의수준보다 작으면 일반적으로 “통계적으로 유의하다”고 판단합니다. 유의수준은 여러 평가항목을 동시에 분석하는 경우, 우연히 통계적으로 유의한 결과가 나타날 가능성이 증가할 수 있습니다. 이런 위험을 줄이기 위해서는 필요에 따라 다중 비교 보정(Multiple comparison adjustment)¹⁷⁾을 적용할 수 있습니다.

■ 유의확률, Significant Probability(p-value)

유의확률은 의료기기 임상시험에서 유효성/성능효과가 통계적으로 의미가 있는지를 판단하는데 활용됩니다. 실제로는 유효성/성능효과가 없다는 전제(가정)하에서, 임상시험 결과가 관찰될 확률을 의미합니다. 만일, 실제로는 유효성/성능효과가 없다고 가정했음에도 불구하고, 임상시험 결과가 관찰될 확률이 크다면(즉, 유의확률이 크다면), 임상시험 결과가 유효성/성능효과가 없다고 할 수 있는 근거가 될 수 있습니다. 하지만, 이 값이 작다면 임상시험 결과가 유효성/성능효과가 없다는 가정에 해당되는 가능성이 낮아진다고 할 수 있습니다. 따라서, 임상시험 결과가 유효성/성능효과가 없다고 할 수 없고, 이를 통계적으로 유의하다고 표현합니다. 일반적으로 유의확률이 크다 혹은 작다의 기준으로 통계적 유의미함을 판단할 수 있는데, 이 기준은 미리 설정해야 하고, 일반적으로 0.05 (유의수준)가 기준으로 사용되는 경우가 많습니다. 예를 들어, 진단 목적의 AUROC 결과를 분석하였을 때 산출된 유의확률이 0.01라고 가정합니다. 유의수준인 0.05보다 작은 경우 AUROC를 통해 진단 예측을 잘한다고 할 수 있습니다. 다만, 유의확률은 결과의 통계적 유의성을 평가하는 지표이며, 임상적 중요성 또는 실제 의료 현장에서 활용가치를 직접적으로 의미하는 것은 아닙니다. 따라서 결과 해석 시에는 평가변수의 특성, 효과의 크기 및 신뢰구간 등을 함께 고려할 필요가 있습니다.

17) 본 안내서에서는 다중비교 보정(다중성 보정이라고도 함)에 대한 설명은 생략하였습니다.

■ 신뢰구간, Confidence Interval(CI)

신뢰구간은 분석결과가 어느 정도의 범위 안에 있을 것으로 예상되는 구간을 나타냅니다. 통계분석은 전체 대상(모집단, population)이 아닌 일부 표본(sample)을 이용하여 수행합니다. 즉, 동일한 임상 시험을 여러번 수행한다고 가정했을 때, 동일한 모집단이라 하더라도 매번 다른 표본이 추출될 것입니다. 따라서, 임상시험들의 결과값도 달라지는 불확실성이 존재할 것입니다. 신뢰구간은 이러한 불확실성을 수치로 표현한 것입니다. 일반적으로는 95% 신뢰구간을 가장 많이 사용합니다. 예를 들어, AUROC 분석 결과, 추정값은 0.88이고, 95% 신뢰구간이 (0.86, 0.89)로 분석되어 표기될 수 있는데, 추정값의 불확실성을 나타내는 범위를 의미로 해석될 수 있습니다.

통계적으로는 동일한 방법으로 임상시험을 여러번 반복했을 때 계산되는 신뢰구간의 약 95%가 실제 참값(진짜 성능값)을 포함하도록 설계된 구간을 의미합니다. 임상시험의 통계 결과를 제시할 때 추정값, p-value와 함께 신뢰구간을 제시하는 것이 중요합니다.

분석 결과, 유의확률은 0.05(즉, 유의수준 5%)보다는 작아야 임상시험 결과가 통계적으로 유의미하고 신뢰구간은 성능의 불확실성 범위로 신뢰구간이 좁을수록 불확실성이 작다고 할 수 있습니다.

■ 중도탈락, Drop-out

중도탈락은 임상시험에 참여한 대상자가 연구 종료 전에 참여를 중단하거나 필요한 자료를 제공하지 못해 최종분석에 포함되지 않은 경우를 의미합니다. 중도탈락은 연구 참여 동의를 철회하거나, 추적 관찰 기간 중 연락이 두절된 경우, 부작용으로 인해 연구 참여를 중단한 경우, 예정된 시점에 평가를 받지 못한 경우, 임상시험 계획서 기준에 따라 제외된 경우 등에서 발생할 수 있습니다.

중도탈락이 많아지면 임상시험 분석 결과의 신뢰성이 낮아질 수 있습니다. 따라서, 임상시험 설계 시 예상중도탈락을 고려하여 충분한 대상자를 모집하고, 중도탈락의 원인을 기록 관리하여, 최종 분석 시 검정력이 손실되지 않는 대상자수로 분석할 수 있도록 고려해야 합니다.

예를 들어, 임상시험 설계 시, 목표하는 효과크기, 참고치, 유의수준 및 검정력을 가정하여 100명의 대상자가 산출되었을 경우, 중도탈락률을 15%로 가정했다면, 최종 모집 대상자수는 $100/(1-0.15) = 117.65$ 로 계산되어 최소 118명을 모집하는 설계가 필요하게 됩니다. 즉, 118명을 등록하면 약 15%인 18명 정도가 중도탈락하더라도 최종적으로 약 100명이 분석에 포함되어 설계했던 충분한 검정력 하에서 분석할 수 있게 되는 것을 의미합니다.

■ 편향, Bias

편향은 임상시험의 결과가 실제와 다르게 나타내도록 영향을 주는 체계적 오류(systematic error)를 의미합니다. 편향이 발생하면 유효성/성능효과나 안전성이 실제보다 좋거나 나쁘게 평가될 수 있어 신뢰할 수 있는 결과를 얻기 위해서는 편향을 최소화하는 것이 중요합니다. 예를 들어, 임상시험에 특정 연령대만 시험대상자로 참여하는 경우나 혹은 연구자가 어떤 시험 대상자가 시험군인지를 인지하고 있는 경우 편향의 발생 위험이 있다고 할 수 있습니다. 이를 최소화하기 위해서는 무작위배정과 눈가림등의 방법이 있으며, 그 외 분석 시 통계적인 보정방법 등이 있습니다.

■ 단일 눈가림, Single-blind

단일 눈가림법은 연구자나 의료진은 배정군을 알지만 임상시험대상자는 모르게 하는 것이며 어떤 경우에는 그 반대의 경우도 있을 수 있습니다. 예를 들어, 새로운 의료제품과 기존 의료제품을 비교하는 임상시험에서, 시험대상자는 자신이 어느 기기를 사용하는지 알 수 없도록 하여 선입견이나 기대감이 결과에 영향을 미치지 않도록 할 수 있습니다.

■ 이중 눈가림, Double-blind

이중 눈가림법은 임상시험대상자, 연구자 중 어느 누구도 임상시험대상자가 어떤 군에 배정되었는지를 모르게 하는 것입니다. 이중눈가림을 통해 시험결과가 특정 방향으로 치우치는 것을 방지하여 공정하고 신뢰할 수 있는 결과를 얻을 수 있습니다.

■ 단순 무작위배정, Simple randomization

대조군과의 비교로 설계된 시험의 경우 시험군 또는 대조군이 배정된 시험대상자/검체의 일반적 정보, 질병 특성 정보의 분포가 균등히 배분될 가능성을 높이는 효과가 있습니다.

■ 블록 무작위배정, Block randomization

일정 수의 시험대상자/검체를 블록으로 설정하여 한 블록 내에 동일한 수로 시험군과 대조군을 배정하는 방법입니다. 이는 모집되는 일련의 시험대상자/검체의 순서에서 질병 관련 특성의 분포가 상이하다 하더라도 각 군에 유사한 배정 가능성을 지니는 장점을 가집니다. 이때 블록의 크기는 다양하게 하여 블록의 크기를 예상할 수 없도록 장치를 마련하는 것이 권장됩니다. 예컨대 시험군과 대조군의 두 개 군에 대해 블록 크기를 4로 할 경우에는 각 두명씩 시험군과 대조군을 배정받도록 하는 것입니다. 블록 무작위배정을 계획할 경우, 블록 크기는 계획서에 제시하지 않습니다.

임상시험 설계 시, 편향을 최소화하기 위한 설계요소로 무작위배정과 눈가림이 있습니다.

■ 배정된 대로 분석, Intention-To-Treat (ITT)

무작위배정된 모든 임상시험대상자를 배정된 대로 주 분석에 포함시켜야 한다는 원칙이 ITT개념입니다. 임상시험 완료 여부와 관계 없이 분석에 포함하여 임상시험 결과의 객관성과 신뢰성은 높이고 편향을 최소화 할 수 있습니다. 예를 들어, 시험대상자의 중도 탈락은 결과에 영향을 미칠 수 있으므로 이러한 시험대상자 탈락으로 인한 편향을 줄일 수 있고, 실제 의료현장에서의 효과를 보다 잘 반영할 수 있습니다. ‘모든 분석 대상자군(FAS, Full Analysis Set)’은 무작위배정된 모든 임상시험대상자를 분석에 포함시키는 배정된 대로의 분석 원칙을 가장 근접하게 적용할 수 있는 분석대상 집단입니다.

■ 계획서 순응 임상시험대상자군, Per-Protocol Set (PPS)

계획서 순응 임상시험대상자군은 모든 분석 대상자군 중에서 중대한 위반 없이 임상시험계획서에 따라 임상시험을 실시한 임상시험대상자집단을 의미합니다. 계획서 순응 임상시험 대상자군은 임상시험 계획서를 그대로 준수했을 경우에 관찰할 수 있는 치료효과의 과학적 모형을 잘 반영하며 모든 분석 대상자군에 비하여 새로운 치료가 더 유효하다고 판단할 가능성이 높아집니다.

FAS는 임상시험에 배정된 시험대상자를 가능한 한 모두 포함하여 분석하는 방법으로, 실제 임상 환경에서의 효과를 반영하는 데 목적이 있습니다. 반면, PPS는 임상시험 계획서를 충실히 준수하여 시험을 완료한 대상자만을 포함하여 분석하는 방법으로 이상적 조건에서의 효과를 평가합니다. 따라서 FAS는 현실적인 효과를, PPS는 계획대로 수행했을 때의 효과를 평가합니다.

■ 하위군 분석, Subgroup analysis

임상시험에서 고려한 목표 모집단의 전체 자료 중 일부를 대상으로 분석을 수행하는 과정입니다. 확증적 하위군 분석과 탐색적 하위군 분석으로 구분할 수 있으며, 확증적 하위군 분석을 위해서는 사전에 연구계획서에 분석할 하위군을 지정하고, 통계적 유의수준 보정을 고려하여야 합니다.

찾아보기

(A)		비교 연구	27
AUROC	5, 27	비교군	10, 11
		비열등성 시험	27
(F)		(人)	
FAS(Full Analysis Set)	13	시험군	10, 11
(I)		신뢰구간	29
ITT(Intention-To-Treat)	13, 31	(o)	
(P)		우월성 시험	27
PPS(Per-Protocol Set)	13, 31	유의수준	7, 28
(ㄱ)		유의확률	28
검정력	8, 28	유효례수	9
결측치	7, 13	이중 눈가림	30
계획서 순응 임상시험대상자군	31	(x)	
(L)		전향적	18
눈가림	11, 12	중도탈락	7, 13, 29
(C)		(E)	
단순 무작위배정	30	특이도	27
단일군 연구	27	(표)	
단일 눈가림	30	편향	30
대상자수	6, 7, 9	평가변수	4, 6
대조군	10	안전성평가변수	4
(O)		이차평가변수	4
모집단	29	일차평가변수	4
무작위배정	11, 12	탐색평가변수	4
민감도	26	표본	29
(B)		(ㅎ)	
배정된 대로 분석	31	하위군 분석	31
블록 무작위배정	30	효과크기	7
		후향적	18

소기업·연구개발자를 위한 알기 쉬운 제품화 가이드 -의료제품 개발단계의 쉬운 임상통계- [민원인 안내서]

발 행 처 식품의약품안전처 식품의약품안전평가원 사전상담과 제품화지원팀

발 행 일 2026년 9월

발 행 인 강석연

편집위원장 김희성

편 집 위 원 김문신, 박진경, 김선희

문 의 처 (28159) 충북 청주시 흥덕구 오송읍 오송생명2로 187
식품의약품안전처 식품의약품안전평가원 사전상담과
전화 : 043-719-2935
팩스 : 043-719-2910
