



인공지능(AI) 의료기기 임상시험방법 설계 가이드라인(민원인 안내서)

2022. 07. 04.



식품의약품안전처

식품의약품안전평가원
의 료 기 기 심 사 부

지침서·안내서 제·개정 점검표

명칭

인공지능 의료기기 임상시험방법 설계 가이드라인(민원인 안내서)

아래에 해당하는 사항에 체크하여 주시기 바랍니다.

등록대상 여부	<input type="checkbox"/> 이미 등록된 지침서·안내서 중 동일·유사한 내용의 지침서·안내서가 있습니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 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 : _____)	
	<input type="checkbox"/> 법령(법·시행령·시행규칙) 또는 행정규칙(고시·훈령·예규)의 내용을 단순 편집 또는 나열한 것입니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	<input type="checkbox"/> 단순한 사실을 대외적으로 알리는 공고의 내용입니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	<input type="checkbox"/> 일회성 지시·명령에 해당하는 내용입니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	<input type="checkbox"/> 외국 규정을 단순 번역하거나 설명하는 내용입니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	<input type="checkbox"/> 신규 직원 교육을 위해 법령 또는 행정규칙을 알기 쉽게 정리한 자료입니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 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	<input type="checkbox"/> 행정사무의 통일을 기하기 위하여 내부적으로 행정사무의 세부 기준이나 절차를 제시하는 것입니까? (공무원용)	<input type="checkbox"/> 예(☞지침서) <input checked="" type="checkbox"/> 아니오
	<input type="checkbox"/> 민원인들의 이해를 돕기 위하여 법령 또는 행정규칙을 알기 쉽게 설명하거나 특정 민원업무에 대한 행정기관의 대외적인 입장을 기술하는 것입니까? (민원인용)	<input checked="" type="checkbox"/> 예(☞안내서) <input type="checkbox"/> 아니오
기타 확인 사항	<input type="checkbox"/> 상위 법령을 일탈하여 새로운 규제를 신설·강화하거나 민원인을 구속하는 내용이 있습니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	☞ 상기 질문에 '예'라고 답하신 경우 상위법령 일탈 내용을 삭제하시고 지침서·안내서 제·개정 절차를 진행하시기 바랍니다.	
상기 사항에 대하여 확인하였음.		
2022 년 7 월 4 일		
담당자 확 인(부서장)		손 승 호 강 영 규

이 안내서는 인공지능 의료기기의 임상시험방법 설계에 대하여 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, 본 안내서는 2022년 7월 4일 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 문서에 대한 의견이나 문의 사항이 있으면 의료기기심사부 디지털헬스규제 지원과에 문의하시기 바랍니다.

전화번호: 043-719-3942~3949

팩스번호: 043-719-3940

목 차

I. 일반사항	1
1. 배경 및 목적	1
2. 적용범위	2
3. 용어의 정의	3
II. 임상시험방법 설계 고려사항	6
1. 시험 데이터셋 선정	6
가. 시험 데이터셋 선정방법 및 수	6
나. 피험자 동의	7
2. 임상시험 방법	8
가. 참조표준(Reference Standard) 확보	8
나. 임상시험 디자인	10
다. 다기관 임상시험	10
라. 무작위 배정 및 눈가림	11
마. 대조군 선정 및 대조기기	11
바. 임상 의 합의 판정	12
3. 유효성 평가기준 및 평가방법	13
가. 평가변수 설정	13
나. 평가결과의 성공기준	16
III. 참고문헌	17

1. 배경 및 목적

「인공지능 의료기기의 허가·심사 가이드라인(민원인 안내서)」에서 기계학습 가능 의료기기(Machine Learning-enabled Medical Devices, MLMD)의 임상적 유효성 검증을 위하여 전향적 연구와 함께 후향적 연구를 적용할 수 있도록 도입하였다.

후향적 임상시험은 임상시험의 결과를 신속하고 효율적으로 도출할 수 있어 임상시험 기간 단축 및 비용이 절감되는 큰 장점이 있다. 하지만, 후향적 임상시험은 기존의 의료용 데이터를 이용하기 때문에 시험 데이터셋의 수집 방법, 유효성 평가 변수 등 다양한 사항들을 고려하여 임상시험을 설계하여야 한다.

이에 본 가이드라인에서는 기계학습 가능 의료기기의 후향적 임상시험방법 설계 시 고려해야 할 정보를 민원인에 제공하여 민원편의 및 허가·심사 업무의 투명성을 제공하고자 한다.

2. 적용범위

본 가이드라인은 의료용 데이터를 분석하여 질병을 진단 또는 관리하거나 예측하는 기계학습 가능 의료기기의 의료기기임상시험계획승인, 제조·수입허가(임상시험 검토 대상) 또는 기술문서 등 심사(임상시험 자료 심사 대상)에 적용한다.

3. 용어의 정의

가. 검정력

대립가설이 참인 경우 귀무가설을 기각하는 올바른 결정을 내릴 확률

나. 교차설계

한 명의 피험자에게 시험기기 및 대조기기를 모두 적용하는 방법으로 시험기기 우선 적용군의 경우 시험기기를 먼저 사용하고 일정 기간이 지난 후 대조기기를 사용하는 방법

다. 기계학습 가능 의료기기(Machine Learning-enabled Medical Devices, MLMD)

의도된 의료목적 달성을 위해 부분 혹은 전체에서 기계학습을 사용하는 의료기기

라. 눈가림

임상시험자, 평가자, 피험자가 임상시험 내용을 알게 됨에 따라 발생하는 편향(Bias)을 줄이기 위한 방법으로 단일 눈가림, 이중 눈가림 등이 있음

마. 동등성 시험

시험기기의 유효성이 대조기기와 동등함을 확인하는 방법

바. 무작위 배정

선택 편향을 줄이기 위해 확률의 원리에 따라 피험자를 각 처치군에 배정하는 과정

사. 비열등성 시험

시험기기의 유효성이 대조기기보다 열등하지 않음을 확인하는 방법

아. 우월성 시험

대조기기에 대한 시험기기의 우월성을 확인하는 방법

자. 인공지능(Artificial Intelligence)

논리, 결정트리, 기계학습 또는 심화학습을 사용하여 언어 이해, 대상 및 소리 인식, 학습 및 문제해결과 같은 사람의 능력을 모방하는 업무를 수행하기 위한 프로그래밍을 하는 컴퓨터

차. 임상시험심사위원회(Institutional Review Board, IRB)

임상시험 계획서 또는 변경계획서, 피험자로부터 서면동의를 얻기 위해 사용하는 방법이나 제공되는 정보를 검토하고 지속적으로 이를 확인함으로써 임상시험에 참여하는 피험자의 권리·안전·복지를 보호하는 목적으로 시험기관 내에 독립적으로 설치한 상설위원회

카. 시험 데이터셋(Test Dataset)

훈련 이후 ML 모델의 성능 평가(동 가이드라인에서는 임상시험에 한함)에 사용되며 훈련하는 동안 기계학습 훈련 알고리즘에 노출되지 않는 데이터 집합

타. 전향적 임상시험

연구하고자 하는 요인을 미리 설정한 후 일정기간 동안 변화를 추적하는 연구법으로 위험요소가 일으키는 변화를 관찰하는 임상시험

파. 참조표준(Reference Standard)

비교, 평가, 훈련 등의 예상 결과로 사용되는 객관적으로 규정된 기준

하. 평행설계

피험자를 서로 다른 처치군(2집단 또는 3집단 이상)에 배정하는 방법으로 시험군과 대조군으로 나뉜

거. 훈련 데이터셋(Training Dataset)

시험 데이터셋과는 무관한 ML 모델의 훈련에 사용되는 데이터 집합

너. 후향적 임상시험

연구대상자와 직접적으로 접촉하지 않으면서 이루어지는 연구로 피험자 모집 대신 이전의 진료 또는 임상시험을 통해 획득된 피험자의 의료용 데이터(진료기록, 의료영상, 생체신호, 병리검사, 유전정보 등의 시험데이터셋)를 이용하여 의료기기의 안전성·유효성 검증을 위해 실시하는 임상시험

더. 1차 유효성 평가 기준

임상시험의 목적을 정확하게 나타내는 기준

러. 피험자(의료기기법 시행규칙 제20조)

임상시험에 사용되는 시험 데이터셋을 제공하는 사람

1. 시험 데이터셋 선정

가. 시험 데이터셋 선정방법 및 수

‘시험 데이터셋’은 임상시험을 위해 수집하는 의료용 데이터로서 진료 기록 또는 기존의 임상시험 과정 중 발생한 환자 데이터 등이 해당된다.

‘시험 데이터셋’은 품질, 수집방법, 종류 등에 따라 후향적 임상시험의 임상 유효성 평가 결과에 영향을 미칠 수 있다.

따라서, ‘시험 데이터셋’ 선정은 매우 중요한 과정으로서, 임상시험 설계 시에는 선정기준 및 제외기준을 명확히 하고 수립된 기준에 따라 데이터를 선정하여야 한다.

선정기준 및 제외기준은 의료기기의 적응증에 따라 질환군, 질환의 빈도, 성별 등 목표 집단(Target Population)을 반영하여야 한다.

또한, ‘시험 데이터셋’은 의료기기의 개발 과정동안 사용된 훈련 데이터셋과의 독립성이 유지되어야 하고 편향(Bias)이 발생하지 않도록 모집된 데이터 집단에서 무작위 배정을 통해 추출할 것을 권고한다.

‘시험 데이터셋’의 수는 대상 질병, 임상시험의 목적, 임상시험 평가 변수, 검정력 등을 고려하여 후향적 임상시험에 적합한 통계학적 방법에 따라 산출할 수 있다.

이러한 ‘시험 데이터셋’의 수는 1차 유효성 평가 변수의 종류, 임상 결과의 기대치 정도, 비교 대상 및 방법 등에 따라 달라질 수 있으며

적절한 통계가설을 설정한 후 피험자 데이터 수 산출 공식을 적용하여 산정하도록 한다.

나. 피험자 동의

피험자 동의는 의료기기 임상시험 실시기준 등 관련 규정(의료기기법 시행규칙 제24조제1항제4호)에 따라 임상시험을 시작하기 전, 피험자로부터 동의를 받고 이를 문서화하여야 한다. 또한, 피험자 동의서 서식, 피험자 설명서 및 그 밖의 문서화된 정보는 임상시험심사위원회(IRB)의 승인을 받아야 한다.

그러나 기존의 의료용 데이터를 활용하는 후향적 임상시험에서는 피험자 동의를 받는 것이 현실적으로 어렵거나 피험자에게 미치는 위험이 매우 낮을 수 있다.

따라서, 후향적 임상시험에서 피험자 동의를 면제하는 것은 충분한 검토가 필요하며 임상시험심사위원회(IRB)의 승인 여부에 따라 피험자 동의 면제를 고려할 수 있다.

2. 임상시험 방법

가. 참조표준(Reference Standard) 확보

참조표준은 진단·예측하고자 하는 질병의 유무나 특정한 상태를 진단한 결과로서 단일 또는 복수의 검사나 방법으로 확보할 수 있으며 피험자의 임상적 추적관찰 등을 포함하기도 한다.

인공지능 기반 의료기기(소프트웨어)의 후향적 임상시험에 적용할 수 있는 참조표준의 예는 다음과 같다.

1) 질병에 대한 명시적인 판정 기준

- (예 1) 생체신호를 분석하여 고혈압의 정도를 진단 보조하는 의료기기(소프트웨어)인 경우 고혈압 판정 수치를 참조표준으로 설정
- (예 2) 다양한 의료정보를 분석하여 전신성 홍반성 낭창(Systemic lupus erythematosus, SLE)을 진단 보조하는 의료기기(소프트웨어)인 경우 명시적으로 정의된 진단 기준을 참조표준으로 설정

※ 전신성 홍반성 낭창의 진단 기준 : 미국류마티스학회(ACR)에서 제안하는 루푸스 분류기준에 따라 ‘원반모양 홍반’ 등 11개의 항목 중 4개 이상의 항목이 확인 되는 경우

2) 참조표준 의료기기를 이용한 검사결과 또는 질병의 표준검사 방법에 의한 확진 데이터

- (예 1) 흉부 X-ray 영상을 이용하여 폐암을 검출하는 의료기기(소프트웨어)인 경우 CT 또는 MR 영상의 검사결과를 통해 확진된 환자의 의료영상을 참조표준으로 설정
- (예 2) X-ray 영상을 이용하여 유방암을 진단 보조하는 의료기기(소프트웨어)인 경우 조직검사를 통해 확진된 환자의 의료영상을 참조표준으로 설정

3) 임상 전문가 그룹에 의해 확진된 데이터

- (예) MR 영상을 이용하여 뇌질환을 진단 보조하는 의료기기(소프트웨어)로 질병의 명확한 판정기준 또는 기타 검사방법이 없는 경우 임상 전문가 그룹에 의해 확진된 데이터를 참조표준으로 설정

임상 전문가 그룹에 의해 확진된 데이터를 이용하여 참조표준을 구축하는 경우는 ① 질병에 대한 명시적인 판정 기준이나 ② 참조표준 의료기기를 이용한 검사결과 또는 ③ 질병의 표준검사 방법에 의한 확진 데이터가 없을 때 가능하다.

이때, 임상 전문가 그룹에 의하여 참조표준을 구축하는 경우, 의료기기의 적응증과 임상시험 목적 등에 적합한 전공과 경력을 갖춘 복수의 임상가가 참여하여야 한다.

모든 임상가는 동일한 진료 지침을 이용하여 질병을 판독하고, 판독 결과의 의견 불일치가 발생할 경우에는 적절한 합의과정을 통해 해결한다. 또한, 필요한 경우 해당 질환과 관련한 임상 학회나 단체 등의 견해를 반영하여 결정할 수 있다.

다만, 임상시험 시 편향(Bias)을 최소화하기 위하여 참조표준 구축에 참여한 임상가는 임상시험 평가자로 참여하지 않을 것을 권고한다.

□ 의료기기 임상시험계획서 작성 시 기재사항

- ㉠ 참조표준의 유형
- ㉠ 참조표준 구축 시 필요한 조건 및 구축 과정
- ㉠ 참조표준 구축 시 참여하는 임상 전문가의 전공 및 경력
- ㉠ 필요할 경우 참조표준 구축 시 관련 임상 학회 또는 단체의 견해 등 기재

나. 임상시험 디자인

임상시험 디자인은 임상시험의 목적에 맞게 설계하여야 하고 제품의 특성, 유효성 입증방법(우월성, 동등성, 비열등성 평가 등), 유효성 평가기준 등을 고려하여야 한다. 또한, 후향적 임상시험도 임상시험의 목적에 따라 평행설계나 교차설계를 적용할 수 있다.

인공지능 기반 의료기기(소프트웨어)의 후향적 임상시험에 적용할 수 있는 임상시험 디자인의 예는 다음과 같다. 하지만 반드시 아래에 예시에 국한되는 것은 아니다.

1) 의료기기의 진단 유효성 평가를 위한 임상시험 디자인

- (예 1) 시험 데이터셋을 단일군으로 구성하여 참조표준과 의료기기의 판독 결과 값을 비교
- (예 2) 동일한 시험 데이터셋을 시험군과 대조군으로 각각 배정하고 시험군인 의료기기의 판독 결과 값과 대조군인 임상치의 판독 결과 값을 비교

2) 의료기기 사용으로 임상치의 진단을 향상을 평가하기 위한 임상시험 디자인

- (예) 시험 데이터셋을 단일군으로 구성하여 해당 제품 사용 전·후의 임상치의 판독결과 정확도 등을 민감도·특이도, AUC 등으로 비교

다. 다기관 임상시험

다기관 임상시험은 충분한 수의 시험 데이터셋을 모집하거나 임상시험 결과를 일반화하는데 좋은 근거를 제공할 수 있다.

후향적 임상시험에서는 다기관 임상시험이 반드시 필요하지 않으며 임상시험 의뢰자와 참여기관과의 협의 하에 다기관 임상시험 여부를 결정할 수 있다.

라. 무작위 배정 및 눈가림

무작위 배정과 눈가림은 임상시험에서 편향을 최소화하기 위한 중요한 설계 방법이다.

후향적 임상시험에서의 시험 데이터셋은 미리 알려진 피험자 데이터 모집단에서 무작위로 추출하여 선정할 수 있다.

다기관 임상시험 시 무작위 배정으로 임상시험 피험자 데이터를 선정하는 경우, 무작위 배정 절차는 중앙에서 관리하고 각 임상시험 실시기관에 따라 별도로 무작위 배정을 할 수 있다.

또한, 임상시험 설계 시 시험군 또는 대조군 평가에 임상의를 참여한다면 임상의는 임상시험 피험자 데이터의 판독 결과를 알 수 없도록 눈가림을 해야 한다.

□ 의료기기 임상시험계획서 작성 시 기재사항

- ㉠ 편향을 최소화하기 위해 사용한 무작위 배정과 눈가림 방법 기재
- ㉡ 편향을 최소화하기 위해 이외의 다른 방법을 선택한 경우 그 방법 및 근거 기재

마. 대조군 선정 및 대조기기

의료기기의 판독 정확도와 임상의의 판독 정확도를 비교하여 임상 유효성을 평가하고자 하는 경우, 임상의는 대조군 평가자로서 참여할 수 있다. 이때, 대조군 평가에 참여하는 임상의는 참조표준 구축에 참여한 임상 의 전문가 그룹과는 독립적으로 구성할 것을 권고한다.

후향적 임상시험 시 대조기기의 선정은 반드시 필요하지 않으며, 임상시험의 목적에 따라 대조기기의 선정이 필요 한 경우에는 이를 반영하여 임상시험을 설계할 수 있다.

□ 의료기기 임상시험계획서 작성 시 기재사항

- ㉞ 대조군 평가에 임상의를 참여하는 경우 임상의의 선정 조건(전공, 경력, 임상 의 수 등)
- ㉞ 대조기기를 설계하는 경우 비교평가를 하는 의료기기의 허가번호, 품목명, 모델명, 제조회사 등 기재

바. 임상 의 합의 판정

복수의 임상 의가 참조표준을 구축하거나 임상 시험에 참여하는 경우, 질병의 진단 결과는 임상 의들의 합의로 도출할 수 있다.

임상 의들의 합의 방법에는 객관적인 판독 결과를 얻기 위한 절차를 포함하고 개별 판독자의 판단 근거, 불일치 사유, 합의 근거를 명확히 확인할 수 있도록 한다. 이러한 합의 방법은 임상 시험에 참여하는 판독자의 수를 고려하여 업체에서 자율적으로 설정할 수 있다.

예를 들어, 3인의 판독자가 참여하는 경우 2인의 일치된 판독 결과를 채택할 수 있으며, 2인의 판독자가 참여하는 경우는 관련 분야의 경력이 우수한 임상 의가 주도하여 합의할 수 있다.

□ 의료기기 임상시험계획서 작성 시 기재사항

- ㉞ 판독자 수, 임상 의들의 합의 방법 및 절차 등을 자세히 기재

3. 유효성 평가기준 및 평가방법

가. 평가변수 설정

인공지능 기반 의료기기(소프트웨어)의 후향적 임상시험 시 임상 유효성 평가변수로는 참조표준을 이용한 진단의 민감도, 특이도, ROC Curve, AUC 등을 산출할 수 있으며 참조표준이 없는 경우에는 일치율 등을 산출할 수 있다. 필요한 경우(예. 희귀질환 대상의 임상시험 등), 참조표준을 이용하여 유병률에 따른 양성 예측도, 음성 예측도를 추가적으로 산출할 수 있다.

1) 민감도 (Sensitivity)

- 실제로 특정한 질병에 걸린 사람들 중에서 그 질병이 있다고 분류해내는 확률

2) 특이도 (Specificity)

- 실제로 특정한 질병이 없는 사람들 중에서 그 질병이 없다고 분류해내는 확률

3) 양성 예측도 (Positive Predictive Value)

- 특정한 특성을 가지고 있는 것으로 분류된 사람들 가운데 실제로 그 특성을 가지고 있는 사람이 차지하는 분률

4) 음성 예측도 (Negative Predictive Value)

- 특정한 특성을 가지고 있지 않는 것으로 분류된 사람들 가운데 실제로 그 특성을 가지고 있지 않는 사람이 차지하는 분률

5) ROC (Receiver Operating Characteristic) Curve

- 진단검사 결과를 근거로 민감도와 위양성률(1-특이도)을 이용하여 그린 그래프로, 양성과 음성을 구분하는 진단의 성능을 평가할 수 있음

6) AUC (Area Under the Curve)

- ROC(Receiver Operating Characteristic) Curve의 아래 면적으로 진단

정확도를 의미하고 0.5~1.0 사이의 값에서 1에 근접할수록 이상적인 성능이라고 할 수 있음

7) 양성 일치율 (Positive Percent Agreement)

- 양성 결과를 보이는 진단법이 대조 진단법과 양성 진단 결과가 일치하는 비율

8) 음성 일치율 (Negative Percent Agreement)

- 음성 결과를 보이는 진단법이 대조 진단법과 음성 진단 결과가 일치하는 비율

후향적 임상시험에서 1차 유효성 평가변수에 대한 예는 아래와 같다.

1) 소프트웨어 단일군의 민감도, 특이도, AUC 산출

2) 소프트웨어 판독결과의 민감도, 특이도, AUC와 임상의 판독결과의 민감도, 특이도, AUC를 산출하여 각각 비교

3) 임상의의 소프트웨어 사용 전·후의 민감도, 특이도, AUC 변화 비교

4) 참조표준이 없는 경우 소프트웨어 판독결과와 임상의와의 전체 일치율, 양성 일치율, 음성 일치율, Kappa 통계량 산출

민감도와 특이도, 양성 예측도, 음성 예측도 산출 시 참조표준의 결과와 비교하여 다음과 같이 2×2 표 형식으로 산출 할 수 있다.

< 표 1. 민감도, 특이도, 양성 예측도, 음성 예측도 산출 방법 >

결과		참조표준		
		양성	음성	전체
시험 의료기기 (소프트웨어)	양성	진양성(TP)	위양성(FP)	진양성+위양성
	음성	위음성(FN)	진음성(TN)	위음성+진음성
	전체	진양성+위음성	위양성+진음성	n

○ 민감도 (Sensitivity) = $100 \times \text{진양성(TP)} / [\text{진양성(TP)} + \text{위음성(FN)}]$

○ 특이도 (Specificity) = $100 \times \text{진음성(TN)} / [\text{위양성(FP)} + \text{진음성(TN)}]$

- 양성 예측도 (Positive Predictive Value)
= $100 \times \text{진양성(TP)} / [\text{진양성(TP)} + \text{위양성(FP)}]$
- 음성 예측도 (Negative Predictive Value)
= $100 \times \text{진음성(TN)} / [\text{진음성(TN)} + \text{위음성(FN)}]$

임상의 전문가 그룹에 의해 확진된 데이터를 참조표준으로 사용하거나 민감도와 특이도 산출이 어려운 경우 임상의의 판독 결과와의 비교평가를 통해 일치율 등을 산출함으로써 임상 유효성을 확인할 수 있다.

일치율은 진단의 정확도를 의미하는 지표가 아니므로 민감도와 특이도 분석에 비해 낮은 수준의 평가 방법이며 일치율에 대한 적절한 판정 기준이 없으므로 해석에 유의하여야 한다.

민감도와 특이도 산출이 가능할 경우, 일치율을 평가변수로 선정하는 것을 가급적 배제하도록 하고 민감도와 특이도 산출이 어려운 경우에만 일치율을 평가변수로 선정하여 전체 일치율, 양성 일치율, 음성 일치율, Kappa 통계량을 제시하여야 한다.

< 표 2. 일치율 산출 방법 >

결과		비교방법		
		양성	음성	전체
시험 의료기기	양성	a	b	a+b
	음성	c	d	c+d
	전체	a+c	b+d	n

- 양성 일치율 (Positive Percent Agreement) = $100 \times a / (a+c)$ (%)
- 음성 일치율 (Negative Percent Agreement) = $100 \times d / (b+d)$ (%)
- 전체 일치율 (Overall Percent Agreement) = $100 \times (a+d) / n$ (%)
- Kappa 통계량 = $2(ad-bc) / [(a+b)(b+d)(a+c)(c+d)]$
※ Kappa 통계량 : 평가자 사이의 의견이 일치하는 지를 평가하는 통계량

나. 평가결과의 성공기준

임상시험 평가결과에 대한 성공기준은 업체가 자율적으로 설정할 수 있으며 성공기준을 설정한 사유와 그 근거를 제시할 수 있어야 한다.

성공기준에 대한 설정 방법은 다음 3가지의 대조군 유형 중 1가지 이상을 선택하여 설정할 수 있으며 방법은 아래와 같다.

- 1) 신청기기 & 대조기기 : 기허가 받은 의료기기의 성능을 근거로 목표성능 설정
- 2) 신청기기 & 과거대조군(논문) : 논문의 임상 성공기준 등을 인용하여 설정
- 3) 신청기기 & 표준 진단법(의사진료 행위) : 논문 등을 근거로 표준 진단법을 인용하여 설정

설정된 성공기준은 제품의 허가·심사 시 성능, 사용방법에 기재하고, 사용자 매뉴얼에 명기하여 사용자가 확인할 수 있도록 하여야 한다.

1. 인공지능 의료기기의 허가·심사 가이드라인(민원인 안내서), 식품의약품안전처(2022)
2. 의약품 임상시험 통계 가이드라인(민원인 안내서), 식품의약품안전처(2016)
3. 체외진단용 의료기기의 임상적 성능 허가심사 가이드라인(민원인 안내서), 식품의약품안전처(2015)
4. 한 손에 잡히는 임상연구 : 설계와 실행의 모든 것, 임상시험리뷰연구회 옮김(2015)
5. Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests, FDA(2007)
6. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices, FDA(2017)
7. Machine Learning-enabled Medical Devices: Key Terms and Definitions, IMDRF(2022)

[전문가협의체 위원]

소속	직위	성명	비고
삼성서울병원	교수	장동경	학계
세브란스병원	교수	김광준	
서울대학교병원	교수	박창민	
분당서울대학교병원	교수	박상준	
서울성모병원	교수	김현성	
서울아산병원	교수	서준범	
아주대학교	교수	박래웅	
연세대학교	교수	유선국	
성균관대학교	교수	신수용	
경희대학교	교수	이승룡	
공주대학교	교수	최대선	
가천대학교	교수	이강운	
서울대학교	교수	박영석	
카이스트	교수	유희준	
(주)라이프시맨틱스	대표	송승재	산업계
(주)카이랩	대표	배인호	
(주)루닛	이사	장민홍	
(주)루닛	이사	이정인	
(주)뷰노	이사	김현준	
(주)아이알엠	대표	최승욱	
(주)라인웍스	대표	조용현	
(주)디지털 헬스케어 파트너스	대표	최윤섭	
삼성전자(주)	차장	김철신	
삼성메디슨(주)	수석	김동환	
한국디지털병원 수출사업협동조합	상무이사	김태형	조합

인공지능 의료기기 임상시험방법 설계 가이드라인 (민원인 안내서)

발행처 식품의약품안전처 식품의약품안전평가원

발행일 2022년 7월

발행인 서경원

편집위원장 이정림

편집위원 강영규, 손승호, 배영우, 한영민, 김현수, 김병남, 정병수, 조예진

28159

충북 청주시 흥덕구 오송읍 오송생명 2로 187

문의처 식품의약품안전평가원 디지털헬스규제지원과

전화: 043-719-3943

팩스: 043-719-3940

28159 충북 청주시 흥덕구 오송읍 오송생명2로 187
오송보건의료행정타운
식품의약품안전처 식품의약품안전평가원
의료기기심사부 디지털헬스규제지원과
TEL : 043)719-3943 FAX : 043)719-3940
<http://www.mfds.go.kr/medicaldevice>



[부패·공익신고 안내] ※ 신고자 및 신고내용은 보호됩니다.
▶ 식약처 홈페이지 “국민소통 > 신고센터 > 부패·공익신고 상담”코너



식품의약품안전처
식품의약품안전평가원