

임상시험의 통계 원칙

2015. 9.



의약품심사부 의약품심사조정과

< 목 차 >

1. 서론	1
1.1 배경 및 목적	1
1.2 범위 및 방향	1
2. 전반적인 임상 개발에 대한 고려	3
2.1 시험 내용	3
2.1.1 개발 계획	3
2.1.2 확증시험	3
2.1.3 탐색시험	4
2.2 임상시험의 범위	4
2.2.1 모집단	4
2.2.2 일차변수와 이차변수	4
2.2.3 복합변수	5
2.2.4 종합평가변수	6
2.2.5 다중일차변수	6
2.2.6 대리변수	6
2.2.7 범주화된 변수	7
2.3 비뚤림을 피하기 위한 설계	8
2.3.1 눈가림	8
2.3.2 무작위 배정	10
3. 임상시험 설계 시 고려사항	13
3.1. 임상시험 설계	13
3.1.1 병행설계	13
3.1.2 교차설계	13
3.1.3 요인설계	14

3.2	다기관 임상시험	15
3.3	비교유형	16
3.3.1	우월성 평가시험	16
3.3.2	동등성 평가시험 혹은 비열등성 평가시험	17
3.3.3	용량-반응관계 평가 임상시험	18
3.4	집단축차설계	19
3.5	목표한 시험대상자 수	19
3.6	자료 수집과 처리	20
4.	임상시험 시 고려사항	21
4.1	임상시험 모니터링과 중간 분석	21
4.2	선정기준 및 제외기준의 변경	22
4.3	임상시험대상자 확보율	22
4.4	목표한 시험대상자 수 조정	22
4.5	중간분석 및 조기 종료	23
4.6	독립 자료 모니터링 위원회의 역할	25
5.	자료 분석 시 고려사항	26
5.1	사전에 구체화된 분석계획	26
5.2	분석 대상군	26
5.2.1	무작위 배정된 임상시험대상자군	27
5.2.2	계획서 순응 임상시험대상자군	30
5.2.3	두 분석대상 임상시험대상자군의 의의	30
5.3	결측치 및 이상치	31

5.4	자료변환	32
5.5	추정, 신뢰구간 및 가설검정	33
5.6	유의성과 신뢰수준에 대한 조정	34
5.7	하위집단, 교호작용과 공변량	34
5.8	자료의 무결성 및 컴퓨터 소프트웨어의 적격성	35
6.	안전성과 내약성 평가	36
6.1	평가 범위	36
6.2	평가변수의 선택과 자료수집	36
6.3	평가대상 임상시험대상자군 및 자료보고	37
6.4	통계적 평가	38
6.5	통합 요약	39
7.	보고	40
7.1	평가 및 보고	40
7.2	임상 데이터베이스의 요약	42
7.2.1	유효성 자료	43
7.2.2	안전성 자료	43
8.	참고 : 임상시험 통계 용어 해설	44

이 가이드라인은 임상시험의 통계 원칙에 대한 식품의약품안전처의 입장을 기술한 것으로, 대외적으로 법적 효력을 가지는 것이 아님

※ 가이드라인이란 대외적으로 특정한 사안 등에 대하여 식품의약품안전처의 입장을 기술한 것임(식품의약품안전처 지침 등의 관리에 관한 규정(식약처 예규))

※ 이 가이드라인에 대한 의견이나 문의사항이 있을 경우 식품의약품안전평가원 의약품심사부 의약품심사조정과로 문의하시기 바랍니다.

전화번호 : 043-719-2937

팩스번호 : 043-719-2900

I. 서론

1.1 배경 및 목적

의약품의 유효성 및 안전성은 ‘의약품 임상시험 관리기준’(「의약품 등의 안전에 관한 규칙」 [별표4])에 따른 임상시험으로 입증하며, 이러한 기준을 따르는 임상시험의 계획과 분석에서 통계의 역할은 필수적이다. 본 가이드라인은 임상시험에 적용되는 통계학적 방법의 원칙들을 통일하기 위한 것으로 임상시험 개발 전반에서 임상시험의 설계, 수행, 분석, 평가의 방향성을 제시하고, 품목허가를 위한 제출 자료 작성 및 의약품의 안전성 및 유효성을 평가하는 데 도움을 주고자 한다.

본 가이드라인은 ICH(International Conference on Harmonisation)가 1998년 2월 5일 Step 4에 상정한 임상시험의 통계적 원칙(E9)에 근거하여 작성되었다.

1.2 범위 및 방향

임상시험의 계획, 수행 및 통계학적 분석에 관한 세부사항은 임상시험계획서에 명확하고 구체적으로 기술한다. 임상시험계획서에 기술된 대로 임상시험을 진행하는 것과 주요 결과 변수에 대한 분석에 관하여 임상시험계획서에 얼마나 잘 기술되었느냐에 따라 임상시험의 최종 결론에 대한 신뢰성이 영향을 받는다.

확증시험(confirmatory trials)은 일차변수로 유효성 외에 안전성 변수(예를 들어 유해사례, 임상검사치 또는 심전도 측정), 약동학 또는 약력학적 변수(생물학적 동등성 확증시험)를 포함할 수 있다. 이 가이드라인에서 제안한 원칙들은 여러 임상시험의 통합 자료로부터 도출되는 확증적 결과 및 주로 탐색시험에 해당하는 초기 의약품 개발단계의 임상시험에도 적용될 수 있다.

이 가이드라인에서 제안한 원칙들의 대부분은 비뚤림(bias)을 최소화하고 정밀도를 최대화하기 위한 것이다. ‘비뚤림’이란 임상시험의 계획, 수행, 분석 그리고 결과 해석 등의 과정에서 치료 효과의 추정치를 참값에서 벗어나게 만드는 요소들의 계통적인 경향성을 말한다(용어해설 참조). 비뚤림을 작게 하기 위해서는 잠재적

요인을 파악하는 것이 중요하다. 비뚤림은 임상시험으로부터 타당한 결론을 도출하는 과정에 큰 영향을 미칠 수 있다. 비뚤림은 치료군의 배정과 같은 임상시험의 설계 과정 (예: 위험성이 낮은 임상시험대상자들이 한 쪽의 치료에만 몰려서 배정되는 경우) 또는 임상시험 시행이나 분석 과정 (예: 임상시험계획서 위반 사항이나 임상시험 결과 등을 미리 알고서 해당 임상시험대상자를 결과 분석에서 제외시키는 경우) 중 치료 효과의 정확한 평가에 영향을 주는 비뚤림이 발생할 수 있다. 비뚤림은 잘 드러나지 않고, 직접적으로 그 효과를 측정할 수 없는 경우가 많으므로 임상시험의 결과와 주요 결론에 대한 견실성(robustness)을 평가하는 것이 중요하다. 견실성은 가정이 달라지거나 다른 방법으로 자료를 분석하더라도 임상시험의 치료 효과에 대한 주요 결론이 크게 달라지지 않는 것을 의미한다. 임상시험의 결과에서 제시되는 통계적 수치들에 대한 해석을 할 때에는 p-값, 신뢰구간, 추론에 미치는 비뚤림의 잠재적 영향을 고려한다.

임상시험의 설계 및 분석의 주된 방법은 '확률 통계적 방법(Frequentist, 용어해설 참조)'을 기반으로 하기 때문에 이 가이드라인은 가설검정 또는 신뢰구간 등을 말할 때 가장 많이 사용되는 방법을 주로 언급하였다. 그러나 이것은 다른 접근 방법이 적절하지 않다는 것을 의미하는 것은 아니며, 사유가 명확하고 최종 결론이 충분히 견실성이 있는 경우에는 베이지안(Baysian, 용어해설 참조) 방법 또는 다른 접근 방법들도 고려할 수 있다.

2. 전반적인 임상 개발에 대한 고려

2.1 시험 내용

2.1.1 개발 계획 (Development Plan)

신약 개발 시 임상시험의 목적은 일반적으로 약물 개발에 따른 유익성/위해성 상관관계가 허용되는 선에서 의약품의 안전성과 유효성이 나타날 수 있는 용량 범위나 용법 등을 찾아내는 것이다. 유익성을 얻게 되는 특정 인구 집단 및 의약품이 처방될 수 있는 구체적 적응증 등도 명확히 정의할 수 있다.

이러한 일반적인 목적을 달성하기 위해서 구체적인 목적을 가지고 임상시험을 체계적으로 계획 한다. 개발 계획은 임상시험 계획 단계에서 구체화하며, 적절한 의사 결정 시점과 지식의 축적에 따라 수정될 수 있는 유연성이 필요하다. 품목 허가 신청 자료에는 각 임상시험의 개발 계획의 주요 내용과 전체적인 목적을 기술한다.

2.1.2 확증시험 (Confirmatory Trial)

확증시험은 통상 미리 가설이 제시되고 평가되는 적절하게 잘 통제된 임상시험을 말한다. 일반적으로 확증시험은 안전성 또는 유효성에 대한 확실한 증거를 제시하기 위하여 주된 가설은 임상시험의 주요 목적으로부터 도출되고, 미리 정의되며, 임상시험 완료시에 검증이 뒤따른다. 확증시험에서는 연구 대상이 되는 치료 효과의 크기를 적절한 정밀도를 가지고 추정하고, 이것을 임상적 유의성과 연관시키는 것도 중요하다.

확증시험은 주장에 대한 확실한 증거를 제시하여야 하므로 임상시험계획서와 표준작업지침서를 준수하는 것이 특히 중요하다. 그러므로 임상시험 중에 발생 가능한 피할 수 없는 변경 사항들은 설명되고 문서화하며, 그 결과가 어떠한 것이었는지 평가한다. 또한 임상시험 설계에 대한 정당성 및 계획된 분석 원칙과 같은 중요한 통계적 측면 등은 임상시험계획서에 명시하고, 각 임상시험은 제한된 수의 가설들을 다룬다.

2.1.3 탐색시험 (Exploratory Trial)

확증시험은 대부분 이 전에 수행된 탐색시험의 결과에 근거하여 설계된다. 모든 임상시험과 같이 탐색시험의 경우에도 분명하고 정확한 목적을 가져야 한다. 그러나 확증시험과는 달리 탐색시험의 주요 목적이 미리 정의된 가설을 단순히 검정하는 것이 아닐 수 있으며, 축적되는 결과에 따라 분석의 내용이나 관점이 유동적으로 바뀔 수 있도록 하기 위하여 임상시험을 계획하는 과정에서 유연성을 가질 필요가 있다. 탐색시험의 자료 분석은 먼저 자료를 탐색하는 것이며, 이러한 과정을 통해 자료에 근거한 새로운 가설을 형성하게 된다. 따라서 이러한 탐색시험의 결과만으로 유효성을 입증하기에는 한계가 있을 수 있다.

2.2 임상시험의 범위

2.2.1 모집단 (Population)

의약품 개발의 초기 단계에서 연구대상이 되는 임상효과의 발견가능성을 극대화하려는 목적에 따라 임상시험대상자 선정은 큰 영향을 받을 수 있다. 따라서 초기 단계 임상시험에서 시험대상자는 의약품이 궁극적으로 적용될 전체 환자집단 중에서 아주 좁은 범위의 하위 집단이 될 수 있지만, 확증시험에서의 시험대상자는 목표집단(target population)을 보다 더 잘 반영 한다. 따라서 확증시험에서는 치료 효과를 정확히 추정할 수 있을 정도로 충분한 동질성을 유지할 수 있으면, 목표 집단 내에서 선정/제외 기준을 적절히 완화시키는 것이 임상시험의 유용성을 증가시킬 수 있다.

2.2.2 일차변수(Primary Variables)와 이차변수(Secondary Variables)

일차변수(목표치, 일차평가변수)는 임상적 연관성이 가장 높고 임상시험의 주요 목적과 직접적으로 연관되는 증거를 보이는 것으로서 일반적으로 하나의 일차변수를 정한다. 대부분 확증시험의 주요 목적은 유효성에 대한 과학적 증거를 제시하는 것이기 때문에 일차변수는 대부분 유효성 평가변수이다. 안전성/내약성도 때로는 일차변수가 될 수도 있으며, 언제나 중요하게 고려되어야 할 사항이다. 삶의 질

및 의료비와 관련한 측정도 잠재적 일차변수이다.

일차변수를 선정할 때에는 관련 연구 분야에서 선택된 기준 및 표준에 근거하며, 기존 연구 또는 이미 발표된 논문에서 얻어진 경험에 의해 뒷받침되는 믿을 만하고 정당화된 변수를 사용하는 것이 권장된다. 일차변수는 선정/제외 기준에 적합한 임상시험대상자 집단에서 임상적 치료 효과를 타당하고 신뢰성 있게 측정한다는 충분한 근거에 따라 선정 한다. 또한 일차변수는 일반적으로 임상시험대상자 수를 추정하는 데 이용된다(3.5절 참조).

일차변수의 정의 및 선정은 이론적 근거와 함께 임상시험계획서에 기술한다. 눈가림 해제 이후 주요 변수를 재정의 하는 것은 평가하기 어려운 비뚤림을 일으킬 수 있으므로 일반적으로 인정하기 어렵다. 주요 목적으로부터 정의된 임상 효과가 하나 이상의 방법으로 측정될 경우에는 임상시험계획서에 임상적 관련성, 중요성, 객관성 의의, 중요성, 객관성 및 다른 관련된 특성 등에 근거하여 가능하면 하나의 일차변수로 규정한다.

이차변수는 주목적과 관련된 보조적인 측정치이거나 이차적 목적과 관련된 효과의 측정치이다. 이차변수도 임상시험 결과 해석 시 상대적 중요성 및 역할 등을 임상시험계획서에 정의한다.

2.2.3 복합변수(Composite Variables)

만일 주목적과 결부된 다중 측정치로부터 하나의 일차변수를 선택할 수 없는 경우에는 사전에 정의된 알고리즘을 이용하여 다중 측정치를 단일 또는 복합변수로 통합하거나 조합할 수 있다. 실제로 일차변수를 다중적 임상 검사치의 조합으로 설정 할 수 있다(예: 관절염, 정신질환 등에서 사용하는 평가 척도). 이 경우 다중성 문제를 해결하여 제1종 오류를 조정해 줄 필요는 없다. 다중 검사치의 조합 방법을 임상시험계획서에 명시하고, 임상적으로 적절한 유용성의 관점에서 얻어질 결과 척도에 대한 설명도 제시한다. 복합변수가 일차변수로 사용되는 경우에 각 구성 검사치가 타당성과 의미를 가지고 있다면, 종종 구성 검사치 별로 분석된다. 평가 척도가 일차변수로 사용되는 경우에는 내용 타당도(용어해설 참조), 평가자간 신뢰도(용어해설 참조), 평가자내 신뢰도(용어해설 참조), 질병 위중도의 변화 감지를

위한 반응도와 같은 요소들을 고려한다.

2.2.4 종합평가변수(Global Assessment Variables)

전반적인 유효성, 안전성 및 치료의 유용성을 측정하기 위하여 종합평가변수(용어해설 참조)를 설정하기도 한다. 이러한 형태의 변수는 객관적 변수와 환자의 상태나 변화에 대한 연구자의 전체적 소견 등을 통합하는 것이며, 일반적으로 순서화된 범주형 척도이다. 유효성에 대한 종합 평가변수는 신경과 및 정신과 치료 분야 등에서 특히 잘 확립되어 있다.

2.2.5 다중일차변수(Multiple Primary Variables)

각 변수가 치료 효과의 범위를 충분히 반영할 수 있는 측정치인 경우에 이 중 하나 이상의 변수를 주 유효성 변수로 함께 사용하는 것이 권장되기도 한다. 다중일차변수로 계획된 임상시험의 경우, 해석방법을 면밀하게 기술하여야 한다. 시험목적에 영향을 미치는 변수가 하나인지 일부인지 또는 전체 변수인지를 확실히 설명한다. 주된 가설 또는 시험 대상의 가설과 평가변수(예: 평균, %, 분포)들은 미리 정의된 일차변수와 관련하여 명확하게 기술하고, 통계적 추론을 위한 접근 방법도 명백히 기술한다. 다중 비교 시에 발생할 수 있는 문제와 관련하여 제 1종 오류에 대한 효과를 설명하고, 제 1종 오류를 조정하는 방법도 임상시험 계획서에 명시한다. 또한 제시된 일차변수간의 상관성의 범위도 제 1종 오류에 대한 영향을 평가할 때 고려한다. 임상시험의 목적이 설정된 모든 일차변수들의 효과를 증명하는 데 있다면 제 1종 오류를 조절할 필요는 없지만 제 2종 오류나 임상시험대상자 수에 대한 영향을 주의 깊게 고려한다.

2.2.6 대리변수(Surrogate Variables)

실제적으로 임상적 유익성에 대한 직접적인 평가가 현실적이지 않은 경우에 간접적 기준(대리변수)을 고려하여 대리변수가 임상적 유용성에 대한 적절한 예측치라고 알려진 적응증의 경우에는 대리변수를 사용할 수도 있다. 대리변수(용어해설 참조)의 사용 시 고려해야 할 두 가지 중요한 사항이 있다. 첫째, 대리

변수가 연구 대상인 임상적 결과에 대한 실질적 지표가 아닐 수도 있다. 예를 들어 대리변수를 통해 하나의 특정 약리학적 작용기전에 따른 효과는 측정하지만, 치료가 미치는 작용 범위와 궁극적 효과에 대한 완전한 정보를 제공하지는 않는 경우이다. 제시된 대리변수에 대해서는 매우 긍정적 효과를 나타내는 치료가 실제로는 시험대상자의 임상적 결과에 유해한 영향을 주기도 한다. 반대로, 제시된 대리변수에 미치는 측정 가능한 영향 없이 임상적 유의성만을 언급하는 경우도 있다. 둘째로, 대리변수가 이상반응에 대비하여 직접 평가될 수 있는 임상적 유용성에 대한 정량적 측정을 하지 못할 수도 있다. 대리변수의 타당성을 평가하는 통계적 기준이 제안되어 왔으나, 사용 경험은 아직 제한적이다. 실제로 대리변수 증거의 확실성은 ①생물학적 설명 가능성(예: AIDS 발현에 대한 대리변수로서 CD4 세포수) ②임상적 결과에 대한 대리 변수의 예측치를 역학적 연구로 증명하는 것(예, 항고혈압제의 임상시험에서 사망에 대한 대리변수로서 혈압을 이용하는 것), ③임상시험을 통해 대리변수와 임상적 결과에서 나타난 치료 효과의 일치를 나타내는 증거 등(예: 골다공증에 대한 임상시험에서 골 밀도는 이후의 골절률과 밀접하게 연관되어 있으므로, 골 밀도를 대리변수로 사용하는 경우)에 따라 다르다. 임상변수와 대리변수 사이의 관계는 반드시 동일 질환을 치료하는 다른 기전의 의약품에 대해서도 항상 적용될 수 있는 것은 아니다.

2.2.7 범주화된 변수(Categorised Variables)

이분화(Dichotomisation)된 변수를 사용하거나 혹은 다른 연속, 순위 변수를 범주화하는 것이 바람직한 경우가 있다. ‘성공’이나 ‘반응’에 대한 측정 기준은 이분화의 좋은 사례이며, 이런 경우에는 기준 시점에 대한 개선율을 나타내는 연속변수 또는 역치 수준으로 범주화된 순위를 나타내는 순위, 등급, 척도 등의 관점에서 세부적인 정의가 필요하다. ‘확장기 혈압이 90mmHg 이하로 감소됨’은 가장 일반적인 이분화이다. 범주화는 각 범주가 명확한 임상적 의미를 가질 때 가장 유용하다. 시험결과에 따라서 사후에 기준을 정한다면 비뚤림이 발생할 수 있으므로 범주화 기준은 사전에 임상시험계획서에 명시한다. 범주화는 일반적으로 정보의 손실을 초래하여 분석에서 검정력이 떨어지므로 임상시험대상자 수를 계산하는 경우에 이 점을 고려한다.

2.3 비뚤림을 피하기 위한 설계

임상시험에서 비뚤림을 피하기 위한 가장 중요한 설계 방법은 눈가림(blinding)과 무작위배정(Randomisation)이며, 이 두 가지는 허가신청을 위한 비교 임상시험의 일반적 특성이다. 대부분 대조 임상시험들은 이중눈가림을 시행하며, 치료는 무작위배정 계획에 따라 시험대상자를 배정하고 임상시험 실시기관에는 시험대상자의 배정 번호와 치료기간만을 알려준다. 임상시험에 참여한 연구진 중 임상시험대상자를 접하는 사람은 누구도 무작위 배정표는 물론 임상시험대상자가 어떤 시험군에 배정되었는지도 알 수 없도록 해야 한다.

임상시험계획서에 임상시험 진행 중에 발생할 수 있는 계획서 미준수, 조기 중단, 결측치 등 만족할 만한 분석에 방해가 예상되는 불규칙성을 최소화하기 위한 과정을 자세히 기술함으로써 임상시험 설계 단계에서부터 비뚤림을 감소시킬 수 있다.

2.3.1 눈가림(Blinding)

임상시험을 수행하거나 자료를 분석할 때 임상시험대상자가 어느 치료군에 배정되었는지 알게 됨으로써, 치료군에 대한 정보가 임상시험대상자의 선정, 배정, 임상시험대상자의 시험에 대한 태도, 주 평가변수의 평가, 임상시험 중단의 처리, 분석에서 자료를 제외하는 것들에 영향을 미쳐 발생하는 모든 비뚤림의 발생을 최소화하기 위해 눈가림을 실시한다. 눈가림의 주된 목적은 비뚤림의 가능성이 소실될 때까지 치료군을 확인하지 못하도록 막는 것이다.

이중 눈가림법은 임상시험대상자, 연구자 중 어느 누구도 임상시험대상자가 어떤 치료군에 배정되었는지를 모르게 하는 것이다. 임상시험대상자의 자격 요건을 결정하거나 주 평가변수를 산정하거나 임상시험계획서의 순응도를 평가하는 사람도 포함된다. 눈가림 상태는 임상시험의 진행 기간 동안 계속 유지되어야 하고, 자료의 질적 수준이 적정하다고 판단된 경우에만 눈가림을 해제 한다. 치료나 임상적 평가에 관련되지 않은 의뢰자(예: 임상병리사, 분석화학자, 감사관, 심각한 이상반응 보고에 관련된 자)가 눈가림을 해제해야만 하는 경우에는 적절한

표준 눈가림 해제 과정을 따른다.

단일 눈가림법은 연구자나 그 의료진은 배정 치료군을 알지만 임상시험대상자는 모르게 하는 것이며 어떤 경우에는 그 반대의 경우도 있을 수 있다. 공개 임상 시험은 배정 치료군이 모두에게 다 알려져 있는 경우이다.

이중눈가림이 비뚤림을 방지하기 위한 가장 적절한 방법에 해당되며, 치료약은 투여 전이나 투여 기간 동안 어떤 면(외형, 맛 등)에서든 구별되어서는 안된다.

이상적인 이중눈가림을 실현하는 것은 매우 어려울 수 있다. 예를 들어 수술과 약물치료법의 비교처럼 두 치료법의 양상이 완전히 다를 경우, 또는 캡슐을 사용하면 겉으로 보이는 것은 동일해도 제형의 변화가 약동학 및 약력학적 특성을 변화시키므로 두 제형의 생물학적 동등성 확립이 요구되는 경우, 또는 두 치료의 일일 용법이 다른 경우 등이다. 이러한 경우 이중눈가림을 수행하기 위한 한 가지 방법은 이중위약(double dummy, 용어해설 참조)을 사용하는 것이다. 그러나 이중위약 투여 계획은 임상시험대상자의 동기 유발과 순응도에 부정적인 효과를 미칠 수도 있다. 또한 위장 수술(dummy operative procedure)을 수반하는 등의 윤리적 문제도 이중눈가림을 어렵게 한다. 그럼에도 불구하고 이러한 어려움을 극복하기 위해서 집중적인 노력을 기울인다.

어떤 임상시험에서는 이중 눈가림이 계획되었으나 치료의 효과가 뚜렷하게 나타나서 이중눈가림이 제대로 유지되지 않기도 한다. 이런 경우에는 연구자나 관련 의뢰자에게 특정 검사 결과치(선택된 임상 실험실 검사치) 등을 모르게 함으로써 눈가림이 다소 유지될 수 있다.

이중 눈가림법의 적용이 불가능한 경우에는 단일 눈가림법을 고려한다. 어떤 경우에는 실제적 또는 윤리적인 이유로 공개 임상시험만이 가능할 수도 있다. 단일 눈가림과 공개 임상시험은 유연성을 제공하기는 하지만 연구자가 다음 치료가 어떤 것인지를 알게 되어 임상시험대상자의 선정에 영향을 받지 않도록 하는 것이 무엇보다도 중요하다. 임상시험대상자의 선정은 무작위 배정된 치료를 알기 전에 선행되어야 하며 전화에 의한 무작위배정과 같이 중앙화된 무작위배정 방법을 사용한다. 그리고 임상적 평가는 임상시험대상자 치료에 관여하지 않고 치료군의 배정을 모르는 의료진이 한다. 단일 눈가림 또는 공개 임상시험의

경우에는 비뚤림의 다양한 발생 근원을 최소화 하며, 일차변수가 가능한 한 객관적인 것이 될 수 있도록 한다. 알맞은 눈가림의 정도에 대한 근거는 비뚤림을 최소화하기 위해 선택된 다른 방법과 함께 임상시험계획서에 명시한다.

눈가림을 해제하는 경우는 담당 의사가 임상시험대상자의 치료를 위해 어떤 치료를 받고 있는지를 아는 것이 필수적이라고 여겨질 때에만 고려할 수 있다. 그리고 눈가림 해제의 절차와 시간 등도 문서화한다.

이 가이드라인에서 자료의 눈가림 상태에서의 검토(blind review, 용어해설 참조)는 임상시험 종료(마지막 임상시험대상자에 대한 마지막 관찰)와 눈가림 해제 사이의 기간 동안 이루어지는 자료 검토를 의미한다.

2.3.2 무작위 배정(Randomisation)

무작위 배정은 임상시험대상자 치료군 배정을 보다 신중하게 할 수 있다. 무작위 배정은 임상자료를 분석할 때에도 치료 효과의 정량적 평가에 필요한 통계적 근거를 제공한다. 또한 각 치료군의 예후 인자의 분포가 유사하도록 만들어 주기도 한다. 눈가림과 함께 무작위배정을 사용하면, 치료군 배정을 예측함으로써 발생할 수 있는 임상시험대상자 선정 및 배정에 관한 비뚤림을 제거할 수 있게 된다.

임상시험에서 무작위배정이란 임상시험대상자를 각 치료군에 무작위 배정하는 것이다. 간단한 경우, 이것은 치료의 일련표(또는 교차 시험에서는 치료 순서) 또는 임상시험대상자 번호에 부합되는 코드를 의미한다. 스크리닝 기간이 포함된 일부 임상시험의 경우에는 더 복잡해질 수 있지만, 임상시험대상자 치료군의 사전 배정이나 치료 순서 등은 분명하게 정한다. 임상시험의 설계가 다른 경우에는 무작위배정표를 만들어 내는 절차도 다르다. 무작위배정표는 필요한 경우 재현될 수 있어야 하며, 가능한 한 동일한 난수표를 사용하거나 동일한 컴퓨터 절차와 난수를 이용한다.

단순 무작위배정도 가능하지만 묶음 단위(block)의 무작위배정에서 얻어지는 장점도 있다. 임상시험대상자의 특성이 시간에 따라 변하거나 결과적으로 임상

시험대상자 모집 방식이 수정되는 경우에 묶음 단위의 사용은 치료군 간의 비교성을 증가시킬 수 있으며, 치료군 당 임상시험대상자 수를 거의 같게 해준다. 또한 교차시험의 경우에는 더욱 효율적이고 간편한 해석을 할 수 있는 균형 설계(balanced design)를 가능하게 해준다. 묶음 단위의 크기를 결정할 때는 발생할 수 있는 불균형을 최소화할 만큼 충분히 작게 해야 하지만, 묶음 단위 안에서 배정이 진행됨에 따라 임상시험대상자가 어느 치료군에 배정되는지 예측하지 못하게 할 만큼의 충분한 크기는 갖춘다. 연구자와 다른 스텝들이 묶음 단위의 크기를 알 수 없도록 하기 위하여 묶음 단위의 크기를 2개 이상 사용하여 무작위로 각 묶음이 선택 되도록 하기도 한다.

다기관 임상시험(용어해설 참조)인 경우에 무작위배정 과정은 중앙에서 관리하고 각 임상시험 실시기관에 따라 별도의 무작위배정을 할 수 있다. 즉 각각의 임상시험 실시기관에 따라 층화를 하거나 몇 개의 완전한 묶음 단위를 배정하는 것이다. 일반적으로 투여 전 측정되는 중요한 예후인자들(병의 중증도, 나이, 성별 등)에 의한 층화는 각 층 안에서의 균형된 배정이 가능하도록 해주며, 이것은 소규모의 임상시험에서 더욱 유용하다. 둘 또는 세 개 이상 되는 층화 인자를 사용하는 경우는 거의 없으며, 이런 경우에는 균형있는 배정을 유지하기가 더욱 어렵고 논리적으로도 문제가 된다. 이러한 경우에 임상시험의 나머지 과정들에서 역동적 배정계획(dynamic allocation procedure)을 사용하면 여러 층화 인자들 사이의 균형을 이루는데 도움을 줄 것이다. 무작위배정과 관련하여 층화된 인자들은 이어지는 후속 분석에서도 고려한다.

무작위배정이 될 다음 임상시험대상자는 무작위배정 계획표의 다음 숫자에 대응하는 치료를 항상 받는다. 다음 임상시험대상자에 대한 배정 번호 부여와 치료군 배정은 그 임상시험대상자가 무작위배정될 것이 확실한 경우에만 실시한다. 예측 가능한 무작위배정의 내용을 연구자 등이 미리 알 수 있게 하는 세부 사항은 임상시험계획서에 기록하지 않는다. 무작위배정계획표는 임상시험 의뢰자 또는 임상시험 실시 기간 동안 눈가림의 유지를 보증할 수 있는 독립된 단체가 확실하게 보관 한다. 응급 상황에서 임의의 임상시험대상자에 대해 눈가림을 해제해야 하는 경우, 무작위 배정계획을 열람할 수도 있다. 임상시험계획서는 그에 따른 과정, 필요한 문서, 일련의 치료 및 임상시험대상자에 대한 평가 등 사항을 명시 한다.

역동적 배정이란 배정된 치료군의 현재 균형 상태에 따라서 치료군을 배정하는 방법이다. 그리고 층화된 무작위 배정에서는 임상시험대상자가 속해 있는 층간 또는 층내의 현재 균형 상태에 따라서 치료군을 배정한다. 결정론적인 역동적 배정과정은 지양되어야 하며, 각 치료 배정에 무작위배정의 적정 요소가 통합되어야 한다. 또한 임상시험의 이중눈가림 상태를 유지하기 위한 최선의 노력을 기울인다. 예를 들어 임상시험대상자의 치료군 배정에 대한 정보는 대개 전화를 통해 역동적 배정을 통제하고 있는 중앙 임상시험기관에만 한정하도록 하는 것이다. 중앙 임상시험기관에서는 이렇게 함으로써 임상시험의 선정 기준에 대한 재점검을 할 수 있으며, 임상시험에의 참여 유무를 결정하고 특정 형태의 다기관 임상시험에서 의미를 갖는 특성들에 대한 점검이 가능하다. 이중 눈가림을 위해서는 일반적으로 임상시험용 의약품의 사전 포장 및 라벨링이 필요하지만, 그들의 사용 순서가 반드시 순차적인 것은 아니다. 임상시험기관에서 일하는 다른 사람들이 치료군의 배정표를 알지 못하도록 적정한 컴퓨터 알고리즘을 사용하는 것이 바람직하다. 역동적 배정법을 사용하여야 할 때는 연구(행정) 지원에 따른 복잡성과 분석에 미치는 영향을 세심하게 평가한다.

3. 임상시험 설계 시 고려사항

3.1. 임상시험 설계

3.1.1 병행설계(Parallel Group Design)

확증시험의 가장 일반적인 임상시험 설계는 병행설계로서 이 설계는 임상시험 대상자가 2 개 이상의 군 중 한 군에 무작위 배정되어 각기 다른 치료를 받는 것이다. 이때 치료는 단일용량 혹은 다용량의 시험약과 대조치료를 말하며, 대조 치료에는 위약 혹은 치료효과가 있는 비교약물이 포함된다. 이 설계의 기초를 이루는 가정은 다른 설계의 경우보다는 덜 복잡하다. 그럼에도 불구하고 다른 설계와 마찬가지로 분석과 해석을 까다롭게 하는 추가적인 요소가 있을 수 있다. 예를 들면 공변량, 시간에 따른 반복 측정, 계획에 고려되는 요인간의 상호 작용, 계획서 미준수, 중도 탈락과 참여 중지 등이 있다.

3.1.2 교차설계(Crossover Design)

교차설계에서 임상시험대상자는 2 가지 이상의 치료에 순차적으로 무작위 배정되어 치료 간 비교시 임상시험대상자 자신이 대조군 역할을 하는 것이다. 이러한 방법은 특정 수준의 통계적 검정력을 확보하는데 필요한 연구대상자수를 현저히 줄일 수 있다. 가장 단순한 2×2 교차설계에서는 각각의 임상시험대상자가 두 연속적인 치료기간에 무작위 배정순서에 따라서 두 치료 각각을 받게 되는데, 이 때 두 연속된 치료 기간이 주로 소실기간으로 구분된다. 가장 일반적으로는 n 기간 중 $n(\geq 2)$ 치료를 비교하는 것인데, 이 때 각각의 임상시험대상자는 n 개의 치료를 모두 받는다. 여기에는 각각의 임상시험대상자가 n 치료 중 일부를 받거나 혹은 여러 치료가 한 임상시험대상자에게 반복 투여되는 등의 설계상의 다양성이 존재할 수 있다.

교차설계는 그 결과의 타당성을 감소시키는 문제점을 안고 있다. 주된 문제점은 잔류효과 즉, 연속되는 치료 기간에서 치료의 잔류 영향이다. 부가적인 모델에서 부적절한 잔류효과는 직접적인 치료효과 비교에 비뚤림을 일으킬 수 있다. 2×2

설계에서 잔류효과는 치료와 치료기간 간의 상호작용(용어해설 참조)과 통계적으로 구분될 수 없으며, 각 효과에 대한 상호 대조가 '시험대상자 간'에 있기 때문에 검정력이 약하다. 이러한 문제는 차원 높은 설계로 어느 정도 극복할 수 있지만 완전히 제거할 수는 없다.

그러므로 교차설계를 사용할 때는 잔류효과를 피하는 것이 매우 중요하다. 이는 연구대상 질환과 새로운 치료약물에 대한 적절한 지식을 근거로 연구를 설계함으로써 가능하다. 연구대상이 되는 질환은 만성적이고 안정적이어야 한다. 약물의 효과는 치료기간 중에 충분히 발현될 수 있어야 한다. 휴약기는 약물의 효과가 완전히 소실될 정도로 충분히 길어야 한다. 이러한 조건은 기존의 정보와 자료에 의하여 임상시험 전에 확립되어야 한다.

교차설계에서 주의를 기울여야 하는 또 다른 문제점이 있는데 가장 중요한 것은 중도 탈락 등과 같은 시험대상자 수의 감소로 인하여 분석과 해석이 어렵다는 것이다. 또한, 잔류효과의 가능성은 치료기간 이후에 발생하는 이상반응이 어느 치료에 의한 것인지 분별하기 어렵다는 점으로 이어진다. 교차설계는 일반적으로 중도 탈락하는 임상시험대상자가 작을 것으로 기대되는 제한된 상황에만 적용한다.

동일 약물의 두 제제의 생물학적 동등성을 입증하는데 2×2 교차설계가 일반적으로 사용된다. 건강한 자원자에게 적용하는 경우, 두 투약기간 사이의 휴약기가 충분히 길면 잔류효과는 관련 약동학적 변수에는 영향을 미치지 않는 것으로 보인다. 그러나 개별 치료기에 들어가기 전에 실제로 얻은 자료를 기초로 예를 들어 약물이 검출되지 않는다는 것을 증명함으로써 분석기간 중에 이러한 가정을 확인하는 것이 중요하다.

3.1.3 요인설계(Factorial Design)

요인설계는 치료의 조합을 다양하게 하여 두 개 이상의 치료를 동시에 평가한다. 가장 단순한 예는 2×2 요인설계인데, 예를 들면 A와 B 두 치료방법의 4가지 가능한 조합(A 단독투여군, B 단독투여군, A와 B 동시투여군, A도 B도 없음)중 하나에 임상시험대상자를 무작위 배정하는 방법이다. 많은 경우 요인설계는 A와

B의 상호작용을 검정하는데 사용된다. 시험대상자 수를 주효과 검정에 근거하여 산출하였다면, 상호작용을 알아내기 위한 통계적 검정력은 부족하다. 두 치료 방법에 동시에 이용될 가능성이 있는 경우, 특히 이 설계를 이용하여 A와 B의 상호작용을 검정하려 한다면 통계적 검정력에 대한 고려는 중요하다.

요인설계 적용의 중요한 다른 예는 이전의 임상시험에서 치료 C 및 D 각각의 단일 치료의 유효성이 각 용량마다 확립된 경우에 치료 C와 D 동시 사용에 대한 용량-반응 특성을 확립하는 것이다. C의 용량 m 은 보통 0 용량(위약)을 포함하여 선택되고, D의 용량 n 과 비슷하다. 전체적인 디자인은 $m \times n$ 개의 치료 군으로 구성되고, 각 군은 C와 D의 다른 용량 조합이 투여된다. 반응 면적의 결과 추정값은 임상적 사용을 위한 C와 D의 적절한 조합 용량을 확인하는 데 도움이 된다.

2×2 요인설계는 개개의 치료효과 평가에 필요한 동일한 시험대상자 수로 두 치료의 유효성을 동시에 평가함으로써 임상시험대상자를 효율적으로 이용하는데 쓰이기도 한다. 이러한 전략은 사망률을 주결과변수로 하는 대규모 임상시험에 특히 매우 유용한 것으로 증명되었다. 이러한 설계의 효율성과 타당성은 치료 A와 B간의 상호작용이 없을수록 높아진다. 다시 말하면 A와 B의 주 효과변수에 미치는 영향은 상가적 모델(additive model)을 따르며, 따라서 A의 효과는 실질적으로 B의 존재여부에 관계없이 일정하다는 것이다. 교차설계와 마찬가지로 이러한 조건이 합치된다는 증거가 과거 정보나 자료에 의하여 임상시험 이전에 확립되어야 한다.

3.2 다기관 임상시험(Multicenter Trials)

다기관 임상시험은 두 가지 주된 이유로 실시된다. 첫째, 다기관 임상시험은 새로운 약물을 더욱 효율적으로 평가하는 인정된 방법이다. 즉 어떤 상황에서는 다기관 임상시험이 정해진 기간 이내에 임상시험의 목적을 충족시킬 만큼 충분한 수의 시험대상자를 모집하는 유일한 현실적인 해결책일 수 있다. 원칙적으로 이런 성격의 다기관 임상시험은 임상시험의 어느 단계에서도 수행될 수 있다. 이 때 기관 당 많은 수의 시험대상자를 몇몇 기관에서 모집하는 경우도 있고, 혹은 희귀 질환의 경우, 한 기관 당 소수의 시험대상자를 다수의 기관에서 모집하는 경우도 있다.

둘째, 다기관 임상시험(다수의 시험자)은 임상시험 결과를 일반화(용어해설 참조)하는 데 더 좋은 근거를 제공하기 위하여 설계될 수도 있다. 일반화의 근거는 더 다양한 인구집단으로부터 임상시험대상자를 등록하고 폭넓은 임상환경에서 약물을 투여함으로써, 장차 약물이 사용될 환경에 더욱 적합한 실험적 상황을 제공할 수 있다. 또한 여러 명의 시험자가 임상시험에 참여함으로써 약물의 유용성에 대한 보다 광범위한 임상적 판단을 제시할 수 있게 된다. 이러한 종류의 임상시험은 의약품 개발의 후기 단계의 확증시험이 될 수도 있고, 여러 시험자와 시험기관이 참여할 수도 있다. 때로는 일반화의 가능성을 더 높이기 위하여 여러 국가에서 수행되는 경우도 있다.

다기관 임상시험의 결과를 의미있게 해석하고 일반화하기 위하여 임상시험계획서에 따른 임상시험 수행과정이 명확하고 모든 임상시험기관에서 유사하여야 한다. 공통의 임상시험계획서를 계획하고 임상시험을 수행하는 것이 중요하며 임상시험 수행 절차는 가능한 완벽하게 표준화한다.

3.3 비교유형

3.3.1 우월성 평가시험(Superiority Trials)

유효성을 과학적으로 입증하기 위해서는 위약대조시험에서 위약에 대한 우월성을 입증하거나, 대조치료약에 대한 우월성을 보이거나, 혹은 용량-반응관계를 증명해야 한다. 이런 종류의 임상시험을 우월성 평가시험(용어해설 참조)이라 하며 특별한 언급이 없는 한 본 가이드라인에서는 우월성 평가시험을 대상으로 하고 있다.

우월성 평가시험에서 유효성이 있는 것으로 나타난 치료약이 존재할 때 위중한 질환의 경우 위약 대조시험은 비윤리적일 수 있다. 이런 경우 대조약으로는 위약이 아닌 기존 치료약의 사용이 고려된다. 위약 대조 혹은 치료약 대조의 적합성은 임상시험에 따라 개별적으로 고려한다.

3.3.2 동등성 평가시험 혹은 비열등성 평가시험(Equivalence or Non-inferiority Trial)

어떤 경우 우월성을 보이려는 목적 없이 시험약을 표준 치료와 비교하기도 한다. 이런 유형은 그 목적에 따라서 2 가지 범주로 구분된다. 즉, 하나는 ‘동등성 평가시험’(용어해설 참조)이고 다른 하나는 ‘비열등성평가시험’(용어해설 참조)이다. 생물학적 동등성 평가시험은 전자에 속한다. 어떤 경우에는, 예를 들면 화합물이 흡수되지 않아서 혈류에 존재하지 않을 때 후발품목이 시판품목과 임상적으로 동등성을 입증하기 위하여 실시되기도 한다.

대부분의 활성대조시험은 시험약의 효과가 활성 대조약보다 나쁘지 않음을 입증하기 위하여 설계되므로 후자에 속한다. 또 다른 경우는 다용량의 시험약을 추천용량과 비교하거나 다용량의 표준치료약과 비교하는 임상시험이 가능하다. 이러한 설계의 목적은 시험약의 용량-반응관계를 보여주는 동시에 활성 대조약과 시험약을 비교하는 것이다.

활성 대조 동등성 평가시험 혹은 비열등성 평가시험에서는 위약을 포함한 임상시험을 수행함으로써 여러 목적을 추구할 수도 있다. 예를 들면, 위약에 대한 우월성을 확립하고, 시험 설계를 입증하며, 동시에 활성 대조약과 유사한 정도의 유효성 및 안전성 정도를 평가하는 것이다. 위약을 시험에 포함하지 않거나 다용량의 신약을 사용하지 않는 활성 대조 동등성 평가시험(혹은 비열등성 평가시험)에는 설계상의 몇 가지 어려움이 잘 알려져 있다. 이는 우월성 시험과는 대조적으로 내적 검증의 측정이 부족한 것과 관련이 있어 외적 검증이 필요하다.

대조치료약은 조심스럽게 선택한다. 관련 적응증에서의 유효성이 적절한 우월성 평가시험에서 분명하게 확립/정량화되었고, 활성 대조시험에서 유사한 효과를 보일 것으로 기대되는 치료약이 예로 일반적으로 이용된다. 이러한 목적으로 새롭게 수행되는 임상시험은 임상적, 통계적 발전을 고려하여 기 수행된 활성 대조약이 임상적으로 명백한 유효성을 보인 우월성 평가시험과 동일하게 설계(일차 변수, 활성 대조약의 용량, 시험대상자 선정기준 등)하는 것이 바람직하다. 다만, 다르게 설정하는 경우 이에 대한 타당성을 제시하여야 한다.

동등성 혹은 비열등성을 입증하기 위하여 설계된 임상시험계획서에는 그 의도를 분명히 기술한다. 임상시험계획서에 동등성 경계를 기술하며, 이 경계는 임상적으로 허용가능한 가장 큰 차이이면서 활성 대조 우월성 평가시험에서 관찰된 차이보다 작다. 활성 대조 동등성 평가시험에서는 동등성 경계의 상한치와 하한치가 모두 필요한 반면, 활성 대조 비열등성 평가시험에서는 경계의 하한치만이 필요하다. 동등성 경계는 임상적으로 입증 가능하다.

통계적 분석은 일반적으로 신뢰구간을 사용한다. 동등성 평가시험에서는 양측 신뢰구간이 이용되며 모든 신뢰구간이 동등성 경계 내에 있는 경우 동등하다고 추정된다. 비열등성 평가시험에서는 단측 구간이 사용된다. 제 1종 오류의 선택은 단측 혹은 양측 검정의 사용과는 별도로 고려한다. 시험대상자 수 산정은 이러한 방법에 근거한다(3.5절 참조).

시험약과 활성대조약 간에 차이가 없다는 귀무가설이 유의하지 않다는 결과를 가지고 동등성 혹은 비열등성의 결론을 내리는 것은 적합하지 않다.

또한 분석 대상군 선택에도 주의한다. 치료군이나 비교군에서 참여중지 혹은 중도 탈락된 임상시험대상자는 치료효과가 나타나지 않은 경우가 많고, 따라서 모든 무작위배정된 임상시험대상자를 이용한 통계분석은 치료효과가 동등한 것으로 비뚤림을 나타낼 수도 있다(5.2.3절 참조).

3.3.3 용량-반응관계 평가 임상시험(Trials to Show Dose-response Relationship)

새로운 시험약의 용량과 반응이 어떠한 관련성을 보이는 지는 모든 개발단계에서 다양한 접근에 의하여 얻고자 하는 정보가 된다. 용량-반응 시험은 많은 연구목적들을 충족시키는데 그 중에서도 유효성의 확정, 용량-반응 곡선의 평가, 적합한 초기 용량의 추정, 개인별 용량 조절을 위한 적정한 용법의 확인, 최대 허용용량의 결정 등에 특히 중요하다. 이러한 목적은 임상시험 중 위약을 포함한 다양한 용량에서 얻은 자료를 이용하여 달성된다. 이러한 이유로, 신뢰구간의 추정 및 도표화 방법 등으로 용량과 반응간의 연관성을 추정하는 것은 유의성 검정법을 사용하는 것만큼이나 중요하다. 가설검정은 용량의 점진적 증가나 용량-반응 곡선의 모양과 관련하여 특정 질문에 맞게 적용될 필요가 있다. 통계분석계획(용어해설 참조)은 임상

시험계획서에 자세히 기술하여야 한다.

3.4 집단축차설계(Group Sequential Designs)

집단축차설계는 중간분석에 이용된다(4.5. 참조). 집단축차설계가 중간분석을 가능하게 하는 유일한 설계는 아니지만 가장 흔하게 적용되는 방법인데 개별 시험대상자의 자료가 이용가능한대로 연속적으로 분석을 실시하는 것보다는 임상시험기간 중 정기적으로 시험대상자 결과를 평가하는 것이 더 현실적이기 때문이다.

3.5 목표한 시험대상자 수(Sample Size)

임상시험에서 목표한 시험대상자 수는 알아내고자 하는 의문에 신뢰성있는 답을 제공할 수 있도록 충분한 수여야 한다. 목표 시험대상자 수는 보통 임상시험의 주요 목적에 따라 정해진다. 표본 크기를 다른 근거에 의해 정한다면 그 이유를 분명히 정당화 한다. 예를 들면 안전성 문제, 또는 중요한 이차변수에 근거한 목표 시험대상자 수는 일차적인 유효성에 따른 시험 대상자 수보다 더 많은 수가 요구될 수도 있다.

적정 시험대상자 수는 일차변수, 검정통계량, 귀무가설, 선택된 용량에서의 대립가설, 옳은 귀무가설을 기각할 확률(제 1종 오류), 틀린 귀무가설을 채택할 확률(제 2종 오류), 치료탈락과 계획서 위반을 처리하는 접근법 등과 같은 사항들을 자세히 기술한다. 어떤 경우에는 사건발생률이 통계적 검정력에 대한 주요 관심이며, 필요한 사건수로부터 시험에서 필요한 시험대상자 수로 외삽하기 위한 가정이 필요한 경우도 있다.

목표 시험대상자 수 계산은 계산에 사용된 추정량(예를 들면 분산, 평균치, 반응률, 치료효과의 차이)과 함께 임상시험계획서에 제시하며, 추정량의 근거도 제시한다. 시험대상자 수 산출 시 전제한 가정이 변함에 따라 목표 시험대상자 수의 민감도가 어떻게 달라지는지를 평가하는 것은 중요하며, 이는 가정으로부터 나온 편차 범위에 적절한 연구대상수의 범위를 제시함으로써 가능하다. 확증

시험에서는 가정은 보통 발표된 자료나 과거 임상시험의 결과에 근거한다. 치료 효과(용어해설 참조)의 차이는 환자의 관리에 임상적 관련성이 있는 최소한의 효과 혹은 새로운 치료의 기대되는 효과에 관한 판단에 근거할 수도 있다. 일반적으로 제 1종 오류의 확률은 5% 혹은 그 이하로 정해져 있거나, 다중성을 고려하여 보정되기도 한다. 제 2종 오류는 일반적으로 10%에서 20%로 정해져 있으나 의뢰자는 특히 시험을 반복하기가 어렵거나 불가능한 경우 제 2종 오류를 실현 가능한 한 낮은 수치로 유지하고자 한다. 제 1종 오류와 제 2종 오류를 일반적인 수준과 달리 정하는 것이 허용되기도 하며, 어떤 경우에는 더욱 선호되기도 한다.

목표 시험대상자 수 계산은 일차 평가분석군에서 필요한 시험대상자 수를 말한다. 만약 이것이 ‘모든 무작위배정된 시험대상자군’이라면, 효과적인 크기의 추정값은 ‘계획서 순응 임상시험대상자군’에 비하여 줄일 필요가 있다. 이는 임상시험에 더 이상 참여하지 않거나 순응도가 낮은 임상시험대상자가 자료분석에 포함됨으로써 치료효과가 희석되는 것을 허용하기 위한 것이다. 치료효과의 변이에 대한 가정은 수정될 필요가 있다.

사건발생률이 기대보다 낮거나 변이가 기대보다 큰 경우 자료를 눈가림 해제 또는 치료군간 비교 없이 목표 시험대상자 수 재산정이 가능하다(4.4절 참조).

3.6 자료 수집과 처리

자료 수집과 수집된 자료를 의뢰자에게 전달하는 방법은 종이로 된 증례기록지, 원거리 모니터링 체계, 의학적 컴퓨터 체계 그리고 전자 전달 등의 다양한 체계를 이용하여 이루어질 수 있다. 수집 정보의 유형과 내용은 임상시험계획서를 충실히 따르고 임상시험 수행 전에 구체적으로 결정 한다. 자료 수집 방법은 계획된 분석을 시행하는데 초점을 두고 임상시험계획서 준수 또는 미준수를 확인하는 데 필요한 관련정보(예를 들면 약물투여와 투여시점 평가)를 포함한다. 결측치는 변수값 ‘0’ 또는 ‘특성 없음’과 구분되어야 한다.

4. 임상시험 시 고려사항

4.1 임상시험 모니터링과 중간 분석

임상시험계획서에 따라 임상시험을 신중하게 수행함으로써 임상시험 결과의 신뢰성을 현저히 향상시킬 수 있다. 임상시험 중에 문제점을 조기에 발견하고 재발을 최소화할 수 있도록 모니터링을 적절히 시행한다.

확증시험의 모니터링은 두 종류가 있다. 한 종류는 임상시험의 질적 수준을 높게 유지하도록 감독하는 것이고 다른 하나는 이중 눈가림법을 해제하고 각 치료군을 비교하는 중간분석이다. 두가지 모니터링 모두 임상시험 관련자들에게 각각 다른 종류의 책임이 부과되고 다른 형태의 임상시험자료와 정보에 접근할 수 있으므로, 잠재적인 통계적 비뚤림과 실행적 비뚤림을 통제하기 위하여 각기 다른 종류의 원칙이 적용될 수 있다.

임상시험의 질적 수준을 감독하는 목적으로서의 첫번째 형태의 모니터링은 계획서의 준수여부, 수행된 임상시험자료가 믿을만한지, 계획된 시험대상자 확보율이 달성되었는지, 임상시험 설계시 적용한 가설은 적절했었는지, 환자가 임상시험에 계속 참여할 수 있도록 하였는지 등을 점검한다(4.2 ~ 4.4절 참조). 이 방법은 치료효과 비교 시 필요한 정보에 접근하지 않아도 되고, 또한 눈가림을 해제하지 않아도 되므로, 제 1종 오류에 영향을 미치지 않는다. 이러한 목적으로 모니터링을 실시하는 것은 임상시험의뢰자의 책임이고, 임상시험의뢰자 또는 임상시험의뢰자가 선정한 독립적인 단체가 수행한다. 이런 형태의 모니터링 기간은 보통 임상시험 기관 선정에서부터 시작하여 마지막 임상시험대상자의 자료를 수집하고 정리할 때까지이다.

다른 임상시험 모니터링(중간분석)은 치료효과를 비교하는 것을 포함한다. 중간 분석은 눈가림을 해제한 상태에서의 치료군 배정(실제 치료 배정과 치료군의 배정)과 비교 치료군에 대한 요약 자료들을 알아야 한다. 이것은 사전(중간분석이 이루어지기 이전)에 미리 임상시험계획서에 포함되어 있어야 하고, 이로 인한 비뚤림이 생기지 않도록 주의하여야 한다.

4.2 선정기준 및 제외기준의 변경

임상시험계획서에서 정의된 선정 및 제외기준은 임상시험대상자 모집기간 동안 구체적으로 표현되고 일정하게 유지한다. 그러나 변경을 해야만 하는 경우도 있다. 예를 들어 장기간의 임상시험에서는 임상시험의 외부 또는 중간분석으로부터 얻은 의학적 지식에 의해 임상시험대상자 선정 및 제외기준이 변경될 수도 있다. 또한 모니터요원에 의해 임상시험대상자를 등록하여 선정하는 과정에서 반복적으로 발생하는 위반사항(violation)이 발견되거나 지나치게 엄격한 기준에 의해 임상시험대상자 확보율이 대단히 낮을 경우에는 변경되기도 한다. 그러나 임상시험대상자 선정 및 제외기준의 변경은 눈가림법을 유지하면서 진행되어야 하고 변경된 임상시험계획서 안에 상세히 서술한다. 변경된 임상시험 계획서에는 효과 추정치의 변경으로 시험대상자 수의 보정, 변경된 선정기준에 따른 층화분석 같은 분석계획의 변경 등을 포함한다.

4.3 임상시험대상자 확보율

환자를 장기간에 걸쳐 모집해야 하는 임상시험의 경우 임상시험대상자 확보율을 모니터한다. 만약 계획보다 눈에 띄게 떨어질 경우에는 임상시험의 검정력을 유지하고 선택 비뚤림이나 임상시험의 질적 수준을 저하시킬 수 있는 다른 요소를 제거하기 위해 그 원인을 밝히고 적절한 조치를 취한다. 다기관 임상시험에서는 이런 고려들이 기관별로 적용되어야 한다.

4.4 목표한 시험대상자 수 조정

장기간의 임상시험에서는 대개 원래의 계획과 시험대상자 수 산출의 근거가 되었던 가정을 점검해 볼 수 있다. 이것은 임상시험이 잠정적이고 불확실한 정보에 기초하여 계획된 경우에 특히 중요하다. 눈가림 상태에서의 자료에 대한 중간점검은 전체적인 반응의 변이양상이나 결과발생률 혹은 생존률 등이 기대했던 것과는 다르다는 것을 밝힐 수도 있다. 이런 경우 적절히 변경된 가정에 기초하여 시험대상자 수를 다시 산출할 수 있고, 이 내용을 반드시 변경된 임상시험계획서와 임상시험보고서에 그 근거를 설명하고 기록으로 남긴다. 제1종 오류와 신뢰구간에

대하여 눈가림 상태와 그 이후의 과정을 유지하기 위하여 취해지는 조치가 있다면 이를 설명한다.

4.5 중간분석 및 조기 종료

중간분석(용어해설 참조)은 공식적으로 임상시험을 종료하기 전의 어떤 시점에서나 유효성 및 안전성에 대해 치료군간의 차이를 비교할 목적으로 실시하는 분석이다. 이러한 중간분석의 횟수나 방법, 결과가 임상시험의 해석에 영향을 미치므로 모든 중간분석은 미리 신중하게 계획되고 임상시험계획서 내에 기술한다. 특별한 상황에 의해 사전에 계획되어 있지 않던 중간분석이 필요하기도 한다. 이런 경우에는 치료효과를 비교할 자료에 대한 눈가림상태를 해제하기 이전에 임상시험계획서를 변경한다. 임상시험 종료여부를 결정할 의도로 중간분석을 계획한 경우에는 통계적 모니터링 계획에 따른 집단 측차 설계를 사용할 수 있다(3.4절 참조). 이러한 중간분석의 목적은 연구대상이 되는 치료의 우월성이 확실하거나, 유의한 치료효과의 차이가 나타날 것 같지 않는 경우, 받아들일 수 없을 정도로 명백한 부작용이 있는 경우 임상시험을 조기 종료하기 위한 것이다. 일반적으로 임상시험을 조기에 종료하고자 할 때 유효성으로 결정하는 경우 안전성으로 결정하는 경우보다 더 보수적인 접근이 필요하다. 임상시험의 계획 및 자료에 대한 모니터링과정에서 다수의 최종평가변수(multiple endpoints)를 다루었다면 다중성도 고려할 필요가 있다.

임상시험계획서에는 중간분석 일정을 기술하고 그렇지 않은 경우, 최소한 제1종 오류의 수준을 유동적으로 적용하는 방법이 사용된다는 식의 중간분석이 사용될 가능성을 기술한다. 첫 번째 중간 분석 전에 임상시험계획서 변경을 통하여 보다 자세한 내용을 기술할 수도 있다. 종료기준과 구체적인 방법은 임상시험계획서나 그 변경 시 명확하게 기술한다. 이러한 내용은 서면화하거나 자료 모니터 위원회(Data Monitoring Committee)의 승인을 얻어야 한다(4.6절 참조). 계획된 분석 절차를 지키지 않았을 경우 임상시험 결과의 신뢰도가 저하될 수 있다. 임상시험계획을 변경할 필요가 있는 경우는 가능한 빨리 변경계획서 안에 그에 따라 바뀌게 될 통계분석방법의 변경을 구체화하고 특히 이러한 변경이 초래할 분석에 대한 영향과 결론을 논의한다. 변경된 방법을 적용함으로써 전체

제 1종 오류의 수준을 일정하게 통제할 수 있어야 한다.

중간분석으로 자료와 그 결과의 눈가림이 해제될 수 있으므로 중간분석은 철저히 기밀로 진행한다. 임상시험 관련자들이 분석 결과를 알게 되면 임상시험에 임하는 태도가 변할 수 있고, 임상시험대상자를 모집하는 과정이나 치료효과를 비교하는 과정에서 비뚤림이 초래될 수 있기 때문에 분석 결과에 대해서는 눈가림 상태를 유지한다. 이러한 원칙은 직접 중간분석을 실시하는 사람들을 제외하고는, 의뢰자에 의해 고용된 임상시험담당자나 직원들, 연구진등 임상시험대상자와 접촉하는 모든 사람들에게 적용한다. 연구자에게는 임상시험의 지속여부에 대한 결정과 임상시험 절차의 변경에 대해서만 알린다.

시험약의 유효성 및 안전성을 확인하고자 하는 대부분의 임상시험은 목표한 시험대상자 수를 다 완료할 때까지 진행한다. 즉 임상시험은 윤리적인 문제가 발생하거나, 더 이상의 환자를 등록하더라도 의도한 검정력을 확보하기 어려울 경우에만 조기 종료한다. 그러나 어떤 경우에는 다른 임상시험을 계획하여야 하는 등의 다양한 이유로 의뢰자가 임상시험 중간분석 결과에 미리 접근할 수도 있다. 또 다른 경우에는 일부의 임상시험에서만 윤리적인 이유로 치료 효과의 비교에 대한 검토가 필요할 수도 있다(생명에 위협을 주는 중대한 결과나 사망에 대한 연구). 어떤 경우든지 중간분석 계획은 치료효과를 비교하기 위한 자료에 대하여 눈가림법을 해제한 접근을 시도하기 이전에 발생할 수 있는 통계적, 실험적 비뚤림을 통제하기 위하여 임상시험계획서나 변경계획서에 기록한다.

대부분의 임상시험, 특히 공중 보건에 큰 의미를 갖는 임상시험의 경우에는 유효성 및 안전성 결과의 모니터에 대한 책임은 권한이 명확히 정해져 있는 독립적인 자료 모니터링 위원회(Independent Data Monitoring Committee, IDMC), 자료 및 안전성 모니터링 위원회(Data and Safety Monitoring Board) 혹은 자료 모니터링 위원회(Data Monitoring Committee)라 불리는 독립적인 위원회에 부여한다.

의뢰자가 유효성 및 안전성 자료에 대한 모니터 역할을 함으로써 눈가림이 해제된 자료에 접근하려고 한다면 임상시험의 무결성을 유지하고 정보를 공유하는

것을 적절히 통제하고 제한하기 위하여 특별한 주의가 요망된다. 의뢰자는 내부 모니터링 위원회가 문서화된 표준작업지침서(Standard Operating Procedure)를 준수하였고 중간분석결과를 포함한 중요한 의사결정회의의 회의록이 유지되었다는 것을 보증하고 문서화 한다.

임상시험의 조기종료 유무에 관계없이 적절하게 계획되지 않은 중간분석은 임상시험의 결과를 비뚤리게 하거나, 도출된 결론에 대한 신뢰를 약화시킬 수 있으므로 그러한 분석은 반드시 피한다. 계획되지 않은 중간분석을 실시하는 경우, 임상시험 결과보고서에 이유와 눈가림이 어느 정도 해제되었는지를 설명하고 발생할 수 있는 비뚤림의 잠재적인 크기 및 결과 해석에 미치는 영향을 평가하여야 한다.

4.6 독립적 자료 모니터링 위원회의 역할

임상시험의뢰자는 임상시험의 진행, 안전성 자료, 주요 유효성 변수를 주기적으로 평가하고, 임상시험의 지속, 변경 또는 종료에 대해 자문을 얻기 위해 독립적 자료 모니터링 위원회(용어해설 참조)를 구성할 수 있다. 독립적 자료 모니터링 위원회는 문서화된 표준작업지침을 구비하고, 중간분석 결과를 포함한 모든 회의 기록을 보존한다. 이 문서들은 임상시험 완료한 후에 검토할 수 있다. 중요한 비교 정보가 노출되는 것을 조정하고 임상시험 정보가 노출됨으로써 발생할 수 있는 역효과로부터 임상시험의 무결성을 보호하기 위해서 독립적 자료 모니터링 위원회의 독립성이 보장되어야 한다. 독립 자료 모니터링 위원회는 임상시험심사위원회 또는 윤리위원회와는 별개의 기구로, 통계학을 포함한 관련 전문 분야의 임상시험 전문가를 포함한다.

시험의뢰자측 대표가 독립 모니터링 위원회에 소속되어 있다면 위원회의 표준작업지침에 이들의 역할을 명확히(예를 들어, 주요 논제에 대한 투표권 여부 등) 정의한다. 의뢰자측 사람이 눈가림이 해제된 정보에 접근할 수 있기 때문에 의뢰기관 내에서 중간분석 결과의 노출을 통제하는 절차를 마련한다.

5. 자료분석 시 고려사항

5.1 사전에 구체화된 분석계획

임상시험을 설계할 때 자료에 대한 최종 통계 분석의 기본 형태는 임상시험 계획서의 통계 부분에 기술한다. 이 통계 부분에는 1차 변수에 대한 확증분석의 주요한 형태와 예상되는 분석상의 문제점을 해결하는 방법들을 포함한다. 탐색 시험의 경우에는 보다 일반적인 원칙과 방향을 기술하는 것이 가능하다.

통계분석 계획은 개별적 첨부 문서의 형태로 작성될 수 있고, 이 문서는 임상 시험계획서에 기술된 것보다 더 전문적이고 상세한 내용을 포함할 수 있다. 즉, 통계 분석 계획에는 일차 및 이차 그리고 기타 변수들에 대한 통계분석을 실시하는 전반적인 과정이 상세히 기술되도록 할 수 있다. 최종 통계 분석 계획은 자료에 대한 눈가림상태에서의 검토 결과에 따라 재검토 또는 필요시 변경할 수 있고 눈가림을 해제하기 이전에 최종 결정한다. 최종 분석계획을 결정하고, 그에 따라 눈가림이 해제될 경우 이러한 내용은 공식적으로 문서화 한다.

자료에 대한 눈가림 상태에서의 검토 결과에 따라 임상시험계획서에 기술된 통계분석에 있어서의 기본 원칙을 변경해야 하는 경우, 변경된 임상시험계획서에 이를 문서화 한다. 그렇지 않다면 자료에 대한 눈가림 상태에서의 검토 결과에서 도출된 제안 사항에 따라 통계분석계획을 갱신하는 것으로도 충분할 수 있다. 어떤 경우이든지 임상시험(변경)계획서에 기술된 분석으로부터 얻어진 결과만이 확증적인 것으로 간주 될 수 있다.

5.2 분석 대상군(Analysis Sets)

주 분석에 포함될 임상시험대상자 집단은 임상시험 계획서의 통계 부분에서 정의한다. 또한 임상시험 개시에 적용된 모든 임상시험대상자의 명단을 문서화 하는 것도 유용할 수 있다. 이러한 문서의 내용은 특정 임상시험의 세부사항에 따라 달라질 수 있으나 최소한 인구 통계학적 자료와 질병상태에 대한 기초 자료는 가능한 한 포함한다.

임상시험에 무작위 배정된 모든 임상시험대상자가 모든 선정 기준을 만족하며 전혀 중도 탈락됨이 없이 모든 임상시험과정이 완벽하게 실시되고 자료가 완전하게 기록되었다면, 이러한 임상시험대상자 집단을 분석에 포함시키는 것은 자명하다. 임상시험을 계획하고 실시할 경우, 가능한 한 이러한 이상적인 상황을 달성하는 것이 목표이지만, 실제로 이렇게 완벽할 수는 없다. 그러므로 임상시험계획서의 통계 부분은 이러한 예상 문제점들이 임상시험대상자와 분석자료에 어떻게 영향을 주는지에 대하여 명시한다. 임상시험계획서에는 다양한 형태의 임상시험계획서 미준수, 중도탈락(용어해설 참조), 결측치 등과 같이 만족할 만한 분석을 실시하는데 영향을 미칠 수 있는 어떠한 불균형의 발생에 대해 이를 최소화하기 위한 절차를 구체적으로 명시한다. 임상시험계획서는 이런 문제가 발생하는 횟수를 줄이고 동시에 자료분석 과정에서 발생하는 문제들을 처리하는 방법을 기술한다. 자료에 대한 눈가림상태에서의 검토과정 중에 임상시험계획서 미준수를 처리할 수 있는 분석방법의 변경이 가능한지 고려한다. 임상시험계획서의 중요한 위반사항에 대하여, 위반사항 발생 시간, 그 원인 및 임상시험 결과에 미치는 영향을 확인하는 것이 바람직하다. 임상시험계획서 미준수, 결측치, 그 외 다른 문제들의 발생 빈도 및 종류는 임상시험결과보고서에서 문서화하고, 이들이 임상시험 결과에 미치는 잠재적인 영향을 기술한다.

분석대상군은 1)비뚤림을 최소화하고 2)제 1종 오류가 커지는 것을 피하는 원칙에 따라 결정한다.

5.2.1 무작위 배정된 임상시험대상자군(Full Analysis Set)

‘배정된 대로 분석’ 원칙(용어해설 참조)은 무작위 배정된 모든 임상시험대상자를 주 분석에 포함시켜야 한다는 것이다. 이 원칙을 따르기 위해서는 무작위 배정된 모든 임상시험대상자에 대한 완전한 추적이 필요하다. 실제로 이 원칙은 다음과 같은 이유로 달성하기 어렵다. 본 가이드라인에서 ‘무작위 배정된 모든 임상시험대상자(용어해설 참조)’란 용어는 무작위 배정된 모든 임상시험대상자를 분석에 포함시키는 배정된 대로의 분석 원칙을 가장 근접하고 완전하게 적용할 수 있는 분석대상 집단을 뜻한다. 분석에서 최초의 무작위배정상태를 그대로 유지하는 것은 비뚤림의 발생을 예방하고 통계적 검정의 근거를 마련하는데 중요하다.

많은 임상시험에서 무작위 배정된 모든 임상시험대상자집단을 사용하는 것은 통계적으로 보수적인 즉, 두 치료군 사이에 치료효과가 차이가 없다는 귀무가설을 기각하기 어려운 방법이다. 또한 다양한 상황에서 이 방법은 실제 환자 진료 상황에서 관찰할 수 있는 치료효과의 실제적인 추정치를 제공하기도 한다.

무작위 배정된 모든 임상시험대상자군에서 분석시 일부의 임상시험대상자를 제외시킬 수도 있는 경우가 제한적이긴 하여도 있을 수 있는데, 주요한 선정 기준을 위반하거나 임상시험용 의약품을 단 한 번이라도 투여 받지 못하거나 무작위 배정 이후 자료가 없는 경우가 그 예이다. 그러나 이러한 경우들은 항상 그 이유가 분명해야 한다.

다음과 같은 경우에만 선정기준을 만족시키지 못하는 임상시험대상자를 비뚤림이 발생하지 않게 하면서도 분석에서 제외시킬 수 있다.

- 1) 선정기준이 무작위 배정이전에 평가하는 경우
- 2) 적절한 선정기준 위반에 대한 발견이 완벽하게 객관적으로 이루어 지는 경우
- 3) 모든 임상시험대상자가 선정기준 위반에 대한 검토를 동등하게 받는 경우 (이러한 조건은 공개 임상시험이거나 또는 이중 눈가림 임상시험이라고 할지라도 선정기준 검토이전에 자료의 눈가림이 해제되면 만족하기 어렵다. 이러한 이유 때문에 자료에 대한 눈가림 상태에서의 검토가 필요하다.)
- 4) 특정한 선정기준을 위반한 것으로 밝혀진 모든 임상시험대상자를 제외한 경우

어떤 경우에는 임상시험용의약품을 한 번이라도 투여받지 않았던 임상시험대상자를 무작위 배정된 모든 임상시험대상자군에서 제외하는 것이 합리적일 수 있다. 그러나 배정된 치료법을 알고 있는 것이 치료를 시작할 것인지 하지 않을 것인지를 결정에 영향을 미치지 않는다면 이들 임상시험대상자를 제외시키지 않고 배정된 대로의 분석 원칙을 지킬 수 있다. 어떤 다른 경우에는 무작위배정 이후 자료를 전혀 얻을 수 없는 임상시험대상자를 무작위배정된 모든 임상시험대상자군으로부터 제외하는 것이 필요할 수도 있다. 어떠한 경우이든지 이들 특정한 환자군을 제외시킴으로서 발생할 수 있는 잠재적인 비뚤림의 가능성을 고려하지 않는다면 완전한 분석이라고 할 수 없다.

무작위 배정된 모든 임상시험대상자를 분석에 사용한다면 무작위 배정 이후에 발생한 임상시험계획서 위반은 자료와 도출된 결론에 영향을 미칠 수 있는데, 이는 특히 이러한 계획서 위반의 발생이 치료군에 따라 다르게 나타날 경우에 문제가 된다. 대부분의 경우에 이들 임상시험대상자로부터 얻어진 자료를 분석에 포함시키는 것이 적절하며 이것이 배정된 대로의 분석원칙과도 일치하는 것이다. 한 번 이상의 약물 투여를 받은 이후에 임상시험대상자가 중도탈락되고 이 시점 이후로 어떤 자료도 얻을 수 없었거나 다른 이유로 더 이상 추적관찰을 할 수 없었다면 특별한 문제가 발생할 수 있다. 왜냐하면 이들 임상시험대상자를 최종 분석집단에 포함시키지 않음으로서 배정된 대로의 분석 원칙을 심각하게 손상시킬 수 있기 때문이다. 어떤 이유로든 추적관찰을 할 수 없었던 임상시험대상자에 대해 그 시점에서 주 결과 변수를 측정하거나, 임상시험계획서에 사전 정의된 평가일정에 따라 이후의 자료를 수집하는 것이 이러한 관점에서 중요하다. 주 결과변수가 사망이거나 중대한 질병상태인 임상시험에서는 이러한 자료수집이 특히 중요하다. 이런 방법으로 자료를 수집하려는 계획을 임상시험계획서 안에 기술한다.

결측치를 처리하는 방법은 마지막으로 관찰된 유효한 자료를 마치 평가시점에서 얻어진 것처럼 간주하거나 복잡한 수학적 모형을 사용하는 것에 이르기까지 다양하다. 배정된 모든 임상시험대상자군의 분석에서 모든 임상시험대상자들에 대한 주 결과변수들의 측정치를 얻기 위해서는 임상시험대상자의 최종 결과에 대한 몇 가지 가정이 필요하거나 보다 단순한 결과, 예를 들어 치료 성공 또는 실패를 선택함으로써 가능하다. 임상시험계획서의 통계 부분에는 이 방법들을 사용할 것과 그 근거에 대한 기술을 명확하게 하고 만일 수학적 모형을 이용한 경우에는 전제가 되는 가정들을 분명히 설명한다. 특히 이런 방법을 사용하는 것 자체가 치료효과의 추정에 비뚤림을 유발할 수 있다면 관련 분석의 건실성을 입증하는 것이 또한 중요하다.

몇몇의 문제점은 사전에 예측할 수 없기 때문에 임상시험 종료시점에서 눈가림 상태로 자료를 검토할 때까지 이러한 불규칙을 다루는 방법들에 대한 고려를 미루어 둘 수도 있는데, 이렇게 할 경우 임상시험계획서에 이를 기술한다.

5.2.2 계획서 순응 임상시험대상자군(Per Protocol Set)

‘계획서 순응(띄어쓰기) 임상시험대상자군’(용어해설 참조)이란 ‘타당한 증례’, ‘유효성 평가군’ 또는 ‘최종 평가가능례’등으로 불리기도 하는데, 이들은 무작위 배정된 모든 임상시험대상자군에서 임상시험계획서에 따라 임상시험을 실시하였던 집단을 의미한다. 이들은 다음과 같은 기준에 따라 정의할 수 있다.

- 1) 미리 정한 최소기간 이상의 치료를 완료한 경우
- 2) 일차 변수의 측정치를 이용할 수 있는 경우
- 3) 선정기준을 포함하여 임상시험계획서의 주요한 기준을 위반하지 않은 경우

특정한 임상시험 상황에 맞게끔 눈가림상태를 해제하기 이전에 환자를 계획서 순응 임상시험대상자군에서 제외하는 이유를 정의하고 문서화 한다. 계획서 순응 임상시험대상자군을 사용하면 분석에서 새로운 치료가 더 유효하다는 것을 검정해 낼 가능성이 높아지며 이는 임상시험계획서를 그대로 준수했을 경우에 관찰할 수 있는 치료효과의 과학적 모형을 잘 반영한다. 그러나 가설 검정 및 치료효과의 추정치는 임상시험의 종료에 따라 통계적으로 보수적일 수도 그렇지 않을 수도 있다. 왜냐하면 임상시험대상자가 임상시험계획서에 잘 순응하는 것이 치료방법이나 최종결과에 따라 영향을 받는 경우 심한 비뚤림이 발생할 수 있기 때문이다.

계획서순응 임상시험대상자군을 규정하기 위해 일부의 임상시험대상자를 추려내는 과정에서 발생하는 문제나 다른 임상시험계획서 위반의 경우들을 밝히고 요약한다. 따라서 이러한 과정에서 예상되는 임상시험계획서의 위반은 치료군 배정의 오류, 병용금지약물의 사용, 낮은 순응도, 추적관찰 실패, 결측치 등이 있을 수 있다. 치료군 별로 이러한 문제들이 발생한 횟수와 시간 등에 대한 양상을 평가하는 것이 유용한 방법이다.

5.2.3 두 분석대상 임상시험대상자군의 의의

어떠한 임상시험대상자군을 사용하더라도 주요한 임상시험의 결과가 달라지지 않는다는 것을 입증할 필요가 있다. 확증시험에서는 무작위 배정된 모든 임상시험 대상자군과 계획서순응 임상시험대상자군을 모두 분석하도록 계획하는 것이

적절한데, 이렇게 함으로써 두 가지 분석방법사이에 차이가 있다면 결과에 대한 고찰 및 해석의 단서가 될 수 있다. 어떤 경우에는 분석 대상 임상시험대상자군을 달리 선택함에 따라 결과가 달라지는 것에 대한 추가적인 탐색적 분석을 계획하는 것이 권장될 수 있다. 만일 무작위 배정된 모든 임상시험대상자군에 대한 분석과 계획서 순응 임상시험대상자군에 대한 분석이 본질적으로 같은 결론을 유도할 수 있는데, 계획서 순응 임상시험대상자군 분석을 통해 상당부분의 임상시험대상자가 분석에서 제외되어야 한다면 임상시험의 전반적인 결론에 약간의 문제가 있을 수도 있음을 염두에 둔다. 우월성 임상시험 및 동등성 또는 비열등성 임상시험에서 이들 무작위 배정된 모든 임상시험대상자에 대한 분석과 계획서 순응 임상시험대상자군의 분석은 서로 다른 의의를 갖고 있다. 우월성 임상시험의 경우 무작위 배정된 모든 임상시험대상자군에 대한 분석이 일부 예외적인 경우를 제외하고는 주된 분석이 되어야 하는데, 이는 계획서 비순응군이 일반적으로 치료효과의 차이를 작게 추정함으로써 계획서 순응임상시험대상자군에 대한 분석에서 관찰할 수 있는 치료효과의 차이를 과장되게 추정하는 것을 상쇄시키기 때문이다. 그러나 동등성 임상시험이나 비열등성 임상시험에서는 무작위배정된 모든 임상시험대상자군에 대한 분석을 실시하는 것이 일반적으로 보수적이지 않으며 따라서 그 의미를 주의 깊게 고려한다.

5.3 결측치(Missing Values) 및 이상치(Outliers)

임상시험에서 결측치는 비뚤림을 발생시키는 잠재적인 원인이 된다. 따라서 자료의 수집 및 관리에 대해 임상시험계획서에서 요구하고 있는 모든 사항을 충족시키기 위하여 다양한 노력이 필요하다. 그러나 실제로 결측치는 항상 발생하기 마련이다. 그럼에도 불구하고 결측치를 처리하는 방법이 합리적이고 특히 이러한 방법들이 임상시험계획서에 미리 정의되어 있다면 임상시험은 타당한 것으로 간주될 수 있다. 이러한 방법에 대한 정의는 자료에 대한 눈가림상태에서의 검토결과에 따라 통계적 분석방법을 갱신함으로써 보다 정밀할 수 있다. 현재까지 이러한 결측치를 처리하는 데 공통적으로 권장되는 방법은 없다. 결측치를 처리하는 방법을 달리함에 따라 분석 결과가 달라지는지의 유무를 검토하는 것이 중요한데, 이는 특히 결측치가 많이 발생한 경우에 유용하다.

이상치의 영향을 탐색적으로 검토하는 과정에도 유사한 방법을 적용할 수 있는데 어느 정도까지는 이상치에 대한 통계적 정의는 작위적일 수 있다. 특정한 변수값을 이상치로 정의하는 것은 통계적 또는 의학적 의미에서 이를 정당화 할 수 있을 경우에 가장 확실한데, 이 경우 임상시험계획서의 의학과관련 부분에서는 이상치의 처리에 관한 적절한 방법을 종종 정의에 두기도 한다. 임상시험계획서 또는 통계분석 부분에 정의되어 있는 이상치에 대한 처리방법이 어느 치료법을 선호하는 쪽으로 사전에 정의하여서는 안 된다. 이들 이상치에 대한 분석에 대한 절차는 눈가림상태에서의 자료 검토과정 중에 적절하게 갱신될 수 있다. 사전에 임상시험계획서에 이상치에 대한 처리절차를 규정하지 않은 경우에는 실제값을 이용한 분석과 이상치에 의한 효과를 제거 또는 감소시킨 최소한 하나이상의 분석이 실시되어야 하고 그 결과를 비교 검토한다.

5.4 자료변환(Data Transformation)

분석을 하기 이전에 주요한 결과변수들을 변환할지 유무는 이전의 임상시험에서 얻어진 유사한 자료의 분석 결과에 따라 임상시험 계획과정 중에 이루어 질 수 있다. 자료변환(예를 들어, 제곱근을 취하거나 로그변환을 하는 것)은 임상시험 계획서에 명시하고 그 근거가 분명해야 하는데 이는 특히 주 결과변수의 경우에 중요하다. 자료변환의 근거가 되는 가정을 만족시켰는지의 유무를 결정할 수 있는 일반적인 원칙은 통계학 교과서를 참조할 수 있다. 특히 특정한 변수들을 변환시키는 방법들이 일련의 임상영역에서 개발되어 왔다. 자료변환을 할 것인지 그리고 어떻게 할 것인지에 대한 결정은 어떤 척도를 사용하여 임상적 해석이 가능한가에 따라 달려있다.

기저치대비 변화량, 기저치로부터의 백분율 변화량, 반복 측정된 자료에서의 곡선하 면적, 또는 두가지 상이한 변수의 비 등과 같은 유도 변수들에 대해서도 유사한 고려를 할 수 있다. 이러한 유도변수들을 사용했을 경우에 임상적인 해석은 주의 깊게 실시되어야 하며 그 유도과정의 타당성을 임상시험계획서에 분명하게 제시한다.

5.5 추정(Estimation), 신뢰구간(Confidence Interval) 및 가설검정(Hypothesis Testing)

임상시험계획서의 통계 부분에는 검정될 가설과 임상시험의 일차 목적을 달성하기 위해 필요한 치료효과의 추정치를 분명히 규정한다. 이러한 분석에 사용되는 통계적 분석 방법은 1차 변수, 그리고 가능하다면 2차 변수에 대하여 기술되어야 하며, 근거가 되는 통계 모형이 있다면 이 또한 분명히 기술한다. 가능한 모든 경우에 치료효과의 추정과 신뢰구간의 계산을 함께 포함하며 그 방법도 명시되어야 한다. 추정치의 정밀도를 높이거나 기저치의 차이를 보정하기 위하여 기저치를 이용한 경우(예를 들어 공분산 분석) 이를 기술한다.

단측 또는 양측검정을 실시했는지를 분명하게 밝히는 것도 중요한데, 이는 특히 사전에 단측검정을 실시하겠다고 한 경우에 중요한 의미를 갖는다. 가설검정이 적절하지 않은 것으로 간주되었다면 통계적 결론에 이르게 한 다른 방법들에 대한 기술한다. 단측 또는 양측검정의 논점은 논란이 많으며 이는 통계문헌에서 쉽게 발견할 수 있다. 양측검정에서 일반적으로 사용되는 제 1종 오류의 절반수준으로 단측검정에서의 제 1종 오류를 정하는 것이 인·허가 관점에서 더 적절하다. 이러한 방법은 두 치료효과의 가능한 차이를 추정하기 위해 일반적으로 이용되는 양측 신뢰구간을 사용하는 것과도 일치한다.

특정한 통계적 방법을 사용하였다면 이는 임상시험의 통계적 고안은 물론 분석대상 변수들에 관한 의학적·통계적 최신지견을 반영하는 것이어야 한다. 분석에 포함될 모든 효과들(예를 들어 분산분석)은 사전에 명시되어야 하고 만일 일차적 자료분석 결과에 따라 방법이 변경된 경우 이에 대한 설명이 뒤따라야 한다. 동일한 내용이 공분산 분석의 경우에도 고려한다. 통계분석방법의 선택에서 1차 및 2차 변수들의 분포에 대한 적절한 고려가 선행되어야 한다. 이러한 선택을 할 때(예를 들어 모수 또는 비모수 분석), 치료효과의 차이에 대한 통계적 추정치는 물론 유의성 검정 및 신뢰구간을 동시에 추정하는 것이 중요함을 염두에 둔다. 1차 변수에 대한 1차 분석은 1차 또는 2차 변수들에 대한 부가적인 분석과는 분명히 구분한다. 임상시험계획서의 통계부분 또는 통계분석 계획은 1차 및 2차 변수가 아닌 다른 변수들을 어떻게 요약하고 어떻게 정리할 것인지에 대한 개요를 포함한다. 이러한 문서에는 여러 임상시험에 걸쳐 일관성을 달성하기 위하여

참고한 자료를 명시 또는 포함하며, 이는 안전성 관련 자료처럼 용어의 표준화나 분석 자료의 표시가 일관될 필요가 있는 경우 특히 중요하다. 이미 알려진 약리학적 파라미터나 임상시험대상자 개개인의 임상시험계획서 순응도 정도 또는 생물학적으로 근거가 있는 기타의 관련 자료를 포함한 모델 접근 방법은 실제 또는 잠재적인 유효성의 크기에 대한 유용한 통찰력을 제공할 수 있으며, 이는 특히 치료효과의 추정과 관련하여 중요하다.

5.6 유의성과 신뢰도에 대한 조정

다중성이 존재할 때, 임상시험 자료의 분석에 대한 보통의 확률론적 접근법은 제 1종 오류의 조정을 필요로 할 수도 있다. 다중성은 다중일차변수(multiple primary variables), 치료법의 다중비교, 경시적 반복 측정 및 중간 분석 등으로부터 발생한다. 가능하다면 다중성을 피하거나 줄이는 것이 권장되며, 여기에는 주요 1차변수의 확정 (다중변수의 경우), 결정적인 치료비교 대조 선택 (다중비교의 경우), 곡선하 면적과 같은 요약 측정의 이용 (반복 측정의 경우) 등이 있다. 확증 분석의 경우, 이러한 종류의 방법들을 적용한 후에도 여전히 남아 있는 도는 다중성의 요소는 임상시험계획서 내에 명시되어야 한다. 요컨대, 제 1종 오류의 조정과 그 구체적인 방법을 항상 고려하며, 만일 이러한 조정이 필요하지 않다고 판단될 경우 어떠한 이유로 그러한 판단을 하게 되었는지를 분석 계획 내에 기술한다.

5.7 하위집단(Subgroups), 교호작용(Interactions)과 공변량(Covariates)

종종 치료와는 상관없는 다른 계통적 요소들이 1차 변수의 결과에 영향을 미칠 수 있다. 예를 들면 연령, 성별과 같은 공변량들이 영향을 미칠 수 있고, 다기관 임상시험의 경우 서로 다른 기관에서 치료를 받는 특정 세부집단 간에 차이가 있을 수도 있다. 어떤 경우에는 공변량 또는 세부집단 효과의 영향을 조정하는 것이 분석 계획의 필수적인 부분이고, 따라서 이러한 내용을 임상시험 계획서에 포함한다. 임상시험을 고안하기 전에 미리 1차 변수의 효과에 영향을 미치는 중요한 공변량 및 기타 인자들에 대해 확인하며, 추정의 정밀도를 개선하고 처리군 간의 균형을 위해 분석에서 이러한 것들을 어떻게 고려할 것인지를 명시한다.

만약 하나 또는 그 이상의 요소들을 사용하여 임상시험대상자를 층화하였다면, 분석에서 이러한 요소들에 대한 영향을 고려한다. 또한 이렇게 조정된 결과의 신빙성이 의심된다면 조정하지 않은 분석값을 주분석 값으로 하고 조정된 분석법을 보조분석 값으로 하는 것이 권장된다. 1차변수의 효과에 미치는 임상시험 실시기관 및 (1차 변수의) 기저치의 영향에 대해서도 특별한 주의를 기울인다. 무작위 배정 후에 측정된 공변량의 영향을 주효과 분석에서 고려하는 것은 바람직하지 않은데, 이는 치료에 따라 공변량의 측정값이 달라질 수 있기 때문이다.

또한 치료 효과 그 자체는 하위집단이나 공변량에 따라 다양한 양상을 나타낼 수 있다. 효과가 연령에 따라서 감소하거나 임상시험대상자의 특정 진단군에서 증가하는 것이 그 예이다. 어떤 경우에는 이러한 교호작용이 미리 예견되거나, 특정한 집단(노인 환자군)에서 특히 중요한 의미를 가질 수 있으므로, 하위집단 분석이나 교호작용을 고려한 분석을 확증분석에서 계획한다. 그러나 대부분의 경우 하위집단 분석이나 교호작용에 대한 분석은 탐색적이다. 이 경우 그 내용을 명백히 기술하며, 대개 관찰된 치료효과가 전반적으로 일정하다고 하는 것을 탐색적으로 확인하는 것이 중요하다. 일반적으로 이러한 분석에서 대개 처음에는 교호작용 효과를 통계적 모형에 삽입하여 분석을 실시하며, 특정한 하위집단이나 미리 정의된 공변량의 특별 층 내에서 부가적인 탐색적 분석을 보조적으로 실시한다. 탐색적 분석일 경우 이러한 분석들의 결과는 조심스럽게 해석되어야 하며, 탐색적 분석의 결과에만 근거한 치료 효과 및 안전성에 대한 결론은 대개 받아들여지기 어렵다.

5.8 자료의 무결성 및 컴퓨터 소프트웨어의 적격성

분석 결과 나타나는 수치에 대한 신뢰도는 자료 관리(자료입력, 저장, 증명, 정정 및 복구) 및 자료의 통계적인 처리를 위해 사용한 방법과 소프트웨어 (내외부적으로 기술된 것 모두)의 질적 수준과 타당성에 달려있다. 그러므로 자료 처리는 철저하고 효과적인 표준작업지침서에 근거하여 수행한다. 자료 처리와 통계 분석에 사용된 컴퓨터 소프트웨어는 신뢰할 수 있어야 하며 적절한 방법을 통해 검증된 것이어야 한다.

6. 안전성과 내약성 평가

6.1 평가 범위

모든 임상시험에서 안전성과 내약성(용어해설 참조)에 대한 평가는 중요한 요소이다. 대부분 초기단계에서는 탐색적 성격을 띠고 독성 발현여부에 주로 관심을 두지만, 후기단계에는 보다 많은 임상시험대상자를 대상으로 약물에 대한 안전성과 내약성을 확립한다. 후기단계의 비교시험은 일반적으로 새로운 잠재적인 유해효과를 비뚤리지 않는 방법으로 탐색하는 중요한 방법인데, 이러한 임상시험으로 새로운 유해효과를 밝히는데 검정력이 낮을 수도 있다.

안전성과 내약성에 대한 우월성 및 동등성 평가를 위한 시험계획을 유사약물이나 임상시험약의 다른 용량과 비교함으로써 수행할 수 있다. 이러한 경우에는 안전성 평가를 위한 확증시험에서 얻어진 관련정보에 근거를 둔다.

6.2 평가변수의 선택과 자료수집

임상시험에서 약물의 안전성과 내약성을 평가하기 위하여 선택된 방법과 측정치들은 관련약물의 유해효과, 전임상시험 및 초기단계 임상시험으로부터 얻은 정보, 특정 약물의 약리학/약동학적 특성에 의해 발생 가능한 결과, 투여경로, 임상시험대상자의 특성, 임상시험 기간 등을 포함한 많은 요인들에 의하여 영향을 받는다. 임상화학 및 혈액학적 실험실적 검사, 활력징후, 그리고 질병, 징후, 증상과 같은 임상적인 이상반응들이 안전성/내약성의 주 자료이다. 심각한 이상반응이 발생하거나 이상반응으로 인하여 치료를 중단한 경우에는 이상반응 증례를 등록한다(ICH E2A와 ICH E3참조).

서로 다른 임상시험으로부터 얻은 자료의 용이한 통합을 위하여 임상시험 시행중에는 일관된 자료 수집과 평가 방법을 이용한다. 특히 공통된 이상반응 사전어를 이용하는 것이 중요하며, 이상반응 사전에는 이상반응 자료를 전신-기관 분류, 권장용어 분류, 유사용어 분류(용어해설 참조) 등 3 가지로 구분하여 요약하고 있다. 권장용어는 이상반응을 요약하는 차원의 용어이며, 동일 전신-기관

분류에 해당하는 권장용어는 자료의 기술적 보고서 함께 보고될 수 있다(ICH M1참조).

6.3 평가대상 임상시험대상자군 및 자료보고

전반적인 안전성 및 내약성 평가를 위한 임상시험대상자군은 일반적으로 적어도 1회 이상 임상시험약을 투여 받은 임상시험대상자들을 의미한다. 안전성 및 내약성에 대한 변수는 임상시험대상자로부터 이상반응의 형태, 심각성의 정도, 발현시점 및 기간을 포함하여 가능한 한 포괄적으로 수집한다(ICH E2B참조). 여성이나 고령자(ICH E7 참조), 중환자, 병용약물을 투여 받는 사람과 같은 특정 소집단은 안전성과 내약성에 대한 추가적인 평가를 필요로 하며, 이와 같은 평가를 위하여 특수대상을 위한 임상시험 가이드라인을 참조 한다(ICH E3 참조).

모든 안전성 및 내약성 변수는 평가 시 주의를 요하고, 임상시험계획서에 폭넓은 접근방법을 제시한다. 모든 이상반응은 치료 내용과의 관련성 여부에 관계없이 보고되어야 하며, 임상시험대상자로 부터 얻어진 모든 자료는 평가 시 설명되어야 한다. 임상검사 변수의 측정단위와 참고치의 범위에 대하여 신중하게 정의 한다. 만일 동일한 임상시험이지만 여러 임상병리실을 이용하여 실험실적 측정을 하는 경우에는 단일화된 평가를 실시할 수 있도록 임상실험실간의 다른 측정치와 참고치 범위를 표준화한다. 독성에 대한 평가 기준 역시 임상시험계획서에 미리 구체화되고 정당화한다.

이상반응의 발생률은 보통 위험에 처한 임상시험대상자 수에 대한 이상반응 발생 임상시험대상자 수의 비율로 표현한다. 이러한 평가방법이 언제나 명확한 것은 아니며, 예를 들어 상황에 따라 임상시험대상자 수나 노출의 정도(예: 임상시험 대상자-년수)가 분모로 사용되기도 한다. 발생률 계산의 목적이 위험성을 추정하는 것이든 처치군간의 비교를 행하는 것이든 간에 임상시험계획서 안에 정의하는 것은 중요하다. 장기간의 임상시험을 계획하여, 임상시험의 조기종료 혹은 사망에 대한 실질적인 비율을 예상하는데 특히 중요하다. 위험성을 과소평가하지 않기 위하여 생존분석법을 고려하여야 하고 이상반응의 누적발생률을 산출한다.

임상시험대상자의 질병에 따라 징후나 증상의 발생수준을 무시할 수 없는 경우 (예: 정신과 임상시험)에는 여러 다른 종류의 이상반응에 대한 위험도를 추정할 때 이를 고려한다. 처리방법으로는 1) 치료전 기초자료와 비교하여 이상반응이 새로이 발생하거나 악화된 경우에만 이상반응을 보고한다는 ‘치료 중 발생한 이상반응 (용어설명 참조)’ 개념을 이용하거나, 2) 임상시험대상자의 기저상태의 징후나 증상의 영향을 감소시키기 위하여, 정도의 이상반응은 무시하거나 반복 방문 시에 관찰되는 경우에만 이상반응을 포함시키는 원칙을 정하는 방법도 적절할 수 있다. 이런 방법들은 임상시험계획서에 설명하고 정당화 한다.

6.4 통계적 평가

안전성 및 내약성에 대한 연구는 다차원적인 문제이다. 약물의 몇몇 유해효과는 보통 예상이 가능하고 특별히 모니터 되지만, 발생 가능한 유해효과의 범위는 매우 넓고 예상치 못했던 새로운 효과도 항상 발생 가능하다. 더욱이 병용금지 약제의 사용과 같은 임상시험계획서 위반에서 오는 이상반응은 비뚤림을 일으킬 수도 있다. 이런 배경으로 볼 때, 약물의 안전성 및 내약성에 대한 통계적 분석 및 평가는 매우 어렵고, 따라서 확증시험으로부터 안전성 및 내약성에 관한 확정적 정보를 얻는다는 것은 일상적이라기보다는 예외에 가깝다.

안전성 및 내약성에 대한 자료는 해석을 용이하게 하기 위해 주로 신뢰구간을 함께 보고할 수 있다. 또한 치료 군별 및 임상시험대상자별 이상반응의 형태를 그래프 형태로 보고하는 것이 유용하다.

연구대상인 주요 이상반응의 차이에 대한 평가를 가능하도록 하거나 안전성 및 내약성 변수들 중 추가적인 주의를 기울여야 할 필요가 있는 차이를 강조하기 위하여 p-값의 계산이 유용할 수 있다. 특히 적절한 방법으로 요약하기 어려운 임상검사치의 경우 유용하다. 임상검사치는 평균치를 비교하는 정량적 분석이나 역치를 기준으로 한 범주형 자료를 대상으로 하는 정성적 분석법의 사용이 권장 된다.

가설검정시 다중성을 고려하여 제 1종 오류를 조절하지만 일반적으로 안전성

평가시 제 2종 오류가 더 큰 관심사이다. 다중성을 보정하지 않는 경우에는 통계적으로 유의한 발견을 하더라도 해석에 매우 주의를 기울인다.

대부분의 연구에서 연구자들은 시험약이 대조약물 및 위약과 비교하여 안전성 및 내약성에 있어서 임상적으로 허용할 수 없는 차이가 없다는 것을 입증하려 한다. 유효성의 비열등성 평가시험 및 동등성 평가시험에서와 마찬가지로 신뢰구간을 이용하는 것이 가설검정보다 선호된다. 이러한 방법으로, 발현빈도가 낮아서 정밀도가 상당히 떨어지는 것을 명확하게 설명한다.

6.5 통합 요약

일반적으로 임상시험용의약품 개발 기간 동안 수행된 일련의 임상시험으로부터 얻어진 자료로 의약품의 안전성을 요약하거나, 특히 품목허가를 받기 위하여 자료를 제출할 때 의약품의 안전성 및 내약성에 관한 자료를 요약하게 된다. 이러한 요약된 자료의 유용성은 자료의 품질 수준이 높고 각각의 임상시험이 적절히 잘 통제되었느냐에 좌우된다.

비록 일반적으로 임상시험 전체를 요약하여 위해성/유익성 평가를 수행하지만, 의약품의 전반적 유용성은 항상 위해성/유익성 사이의 균형의 문제이고, 단일 연구에서도 이러한 점을 깊이 고려한다(7.2.2절 참조).

7. 보고

7.1 평가 및 보고

임상시험계획서에 분석의 주요 원칙을 명시한다. 임상시험 종료 후 각종 자료가 모아지고 예비 검토될 때, 계획된 분석법을 눈가림 상태에서 재검토하는 것이 권장된다. 치료에 대하여 눈가림한 상태에서 분석 전 재검토 시 고려할 사항은 : 1) 분석대상으로부터 일부 자료나 임상시험대상자의 제외 2) 자료변환의 가능성 3) 이상치의 정의 4) 최근 다른 연구에서 밝혀진 중요한 공변량의 분석모델 추가 5) 모수적 또는 비모수적 방법의 사용 등이다. 결정사항은 보고서에 명시하고, 눈가림 상태에서 내린 결정은 비뚤림을 최소화하기 때문에 통계학자들이 무작위 배정 코드를 알고 난 후에 얻어진 결론과는 구분한다. 눈가림이 아닌 상태에서 중간분석에 참여한 통계학자 등은 눈가림 재검토나 통계분석 계획의 조정에 참여하지 않는다. 치료효과가 자료상 명확히 드러나서 눈가림이 제대로 이행되지 못하는 경우 눈가림 재검토에도 통계학자나 관련자는 직접 관여하지 않는다.

자료보고 및 표 등 자세한 사항은 눈가림 재검토 시 완료되어 실제 분석 시에는 임상시험대상자 선정, 자료 선택 및 수정, 자료 요약 및 도표화, 추정 및 가설 검정을 포함한 모든 분석계획을 마련한다. 일단 자료의 타당성에 대한 검토가 완료되면, 미리 정해진 계획에 따라 분석을 수행한다. 이러한 분석이 사전에 작성된 계획에 충실하면 할수록 결과에 대한 신뢰성은 높아진다. 임상시험(변경) 계획서, 또는 자료의 눈가림 재검토에 근거한 통계분석법과 실제로 수행한 분석법에 차이가 있는 경우에는 특별한 주의를 기울인다. 계획된 분석법으로부터 벗어난 내용에 대해서는 보다 구체적인 설명을 한다.

시험에 참가한 모든 임상시험대상자는 분석대상에 포함되었는지 여부에 관계없이 보고서에 설명한다. 또한 분석에서 제외된 모든 이유에 대해서도 문서화한다. ; 배정된 대로 분석 군에는 포함되었으나 계획서 순응 임상시험대상자군에서 제외된 임상시험대상자 및 그 제외사유에 대하여도 언급한다. 유사하게, 분석에 포함된 모든 임상시험대상자에 대한 관련 시점에서의 모든 중요한 변수 측정값에 대하여 설명한다.

임상시험대상자 및 자료의 결측, 투여중지 및 임상시험계획 위반이 주 변수의 주요 분석에 미치는 영향에 대하여 신중하게 고려한다. 추적관찰 실패나 치료중지 또는 중대한 임상시험계획서 위반이 있었던 증례에 대하여 명시하고, 그 사유 및 치료, 결과와의 상관성을 포함한 기술적 분석결과를 포함한다.

기술적 통계는 임상시험 보고서를 구성하는 필수 불가결한 부분이다. 적절한 표와 그래프를 사용하여 주 변수 및 2차 변수, 핵심 예후인자와 인구통계학적 변수의 중요한 특성 등에 대하여 명백하게 기술한다. 임상시험의 목적과 관련한 주 분석 결과는 특히 세심하게 기술한다. 유의성 검정결과는 임계값에 대한 대소의 표시보다는 정확한 p-값(예, $p=0.034$)으로 보고한다.

임상시험의 주요 목적으로 제시된 가설에 대한 해답을 제시하는 것이 임상시험 분석의 주목표이기는 하지만, 눈가림 해제 후 분석결과를 근거하여 새로운 가설이 제기되기도 한다. 이러한 경우에 추가적으로 보다 복잡한 통계분석을 적용할 필요가 있으며 계획서에 예정된 분석결과와 분명히 구별하여 그 결과를 보고서에 제시한다.

분석계획 중 공변량으로서 미리 계획되지 않았으나 몇몇 예후적 중요성을 지닌 기저치 측면에서 치료군간 보이지 않는 불균형이 발생할 수 있다. 불균형을 고려한 추가 분석 결과가 미리 계획된 분석법에 의한 결과와 동일하다는 것을 보여줌으로써 이러한 문제를 해결할 수 있다. 만약 그렇지 않다면 불균형이 결론에 미치는 영향에 대하여 검토한다.

일반적으로 사전에 계획되지 않았던 분석은 최소한으로 한다. 주로 치료효과가 몇몇 다른 요인들에 의하여 다양한 경우에 사전에 계획되지 않았던 분석을 한다. 효과가 특히 좋게 나타난 임상시험대상자 세부집단을 확인하기 위하여 시도될 수 있다. 계획되지 않은 세부집단 분석 결과에 대한 과도한 해석의 위험성은 잘 알려져 있으며(5.7절 참조), 이는 되도록 피한다. 임상시험대상자 세부집단에서 치료효과나 유해성이 없는 경우에도 이와 유사한 과도한 해석 문제가 발생할 수 있으므로, 그러한 가능성은 적절히 평가하고 보고한다.

임상시험 결과의 분석, 해석 및 적절한 결과 제시를 통하여 최종적으로 통계적 판단을 한다. 이를 위하여 임상통계학자는 임상시험보고서에 대하여 책임을 지는 연구진의 일원으로 하고, 최종 보고서를 검토하여 승인하는 업무를 맡는다.

7.2 임상 데이터베이스의 요약

품목허가 신청 시 신약개발 과정 중에 수행된 임상시험을 통하여 보고된 모든 유효성 및 안전성 정보에 대한 전반적 요약 및 종합이 필요하다. 또한 필요한 경우, 연구결과들의 통계적 종합과정을 수반할 수 있다.

요약하는 과정에서 많은 특수한 통계적 관심이 있을 수 있다 : 1) 임상시험 수행과정 동안 처치된 군별 인구통계학적 및 임상적 특성의 기술 2) 관련 임상 비교시험의 결과를 검토하여 유효성에 대한 주 가설을 보장하거나 상쇄되는 정도의 파악 3) 품목허가에 영향을 미치는 잠재적 안전성 문제를 확인하는 모든 임상시험의 통합된 데이터베이스로부터 이용 가능한 안전성 정보의 요약과 발생 가능한 안전성 문제의 확인 등이다. 임상시험 설계 시 일련의 임상시험 특히 그 결과를 통합할 가능성이 있는 임상시험에서는 일관된 정의를 사용하고, 자료를 수집하여, 결과 분석 및 해석이 용이하도록 한다. 약물치료, 병력, 이상반응을 상세히 기록에는 공통된 사건을 사용한다. 주변수 및 이차변수에 대한 공통된 정의는 항상 중요하며, 메타분석(용어해설 참조)에도 필수적이다. 핵심 유효성 변수를 측정하고, 무작위화/등록과 관련된 평가 시점을 조정하고, 임상시험계획서 위반 및 탈락 시험대상자의 처리 및 예후 요소를 정의하는 방식은 특별한 사유가 없는 한 시험기간 동안 일정하게 유지한다.

여러 임상시험의 자료통합을 위하여 사용된 통계적 과정은 상세히 기술한다. 분석에 포함되는 임상시험의 선택과 관련되는 비뚤림의 가능성, 임상시험결과의 동질성, 다양한 변동요인에 대한 적절한 모델화에 주의를 기울인다. 가정 및 임상시험의 선택에 대한 결론의 민감성을 파악한다.

7.2.1 유효성 자료(Efficacy Data)

개별 임상시험은 목적을 충족시킬 만큼 임상시험대상자 수가 충분하여야 한다. 필수적으로 핵심 유효성 질문을 제시하는 일련의 임상시험을 요약하여 추가적인 변수에 관한 정보를 얻을 수 있다. 일련의 시험에서 얻은 주 결과는 보통 추정 및 신뢰구간에 주안점을 둔 표와 그래프를 사용하여 비교 가능한 동일한 형식으로 제시한다. 종종 이러한 추정을 결합하는 메타분석이 유용한데, 치료효과의 크기에 대하여 보다 정밀한 추정을 가능하게 하고, 임상시험의 결과에 대한 완전하고 간략한 보고가 가능하기 때문이다. 예외적인 경우에 메타분석은 전체적인 가설 검정을 통해 유효성에 대한 충분한 근거를 제공하는 가장 최선 또는 유일한 방법일 수 있다. 이러한 목적을 위하여 메타분석을 이용하고자 할 때는 메타분석 자체를 수행하기 위한 전향적인 임상시험계획서를 작성한다.

7.2.2 안전성 자료(Safety Data)

안전성 자료 요약 시 잠재적인 유해성의 모든 징후를 찾기 위하여 철저히 안전성 데이터베이스를 조사하고, 관찰 결과를 지지할 수 있는 징후를 찾기 위하여 추적 관찰하는 것이 중요하다. 의약품에 노출된 인체로부터 발생한 안전성 자료의 조합은 모든 임상시험대상자의 통합된 안전성 자료를 통합하여 연구대상수가 많아지므로 희귀한 이상반응을 발견할 확률이 높아지며 발생률도 근사적으로 추정할 수 있게 된다. 그러나 이러한 데이터베이스로부터 추정된 발생률은 대조군이 없기 때문에 적절한 평가가 어려우므로, 비교시험으로부터 얻은 데이터가 유용할 수 있다. 동일한 대조약(위약 또는 특정 치료효과가 있는 대조약)을 사용한 여러 시험의 결과들은 통합되거나 각 대조약별로 구분하여 보고한다.

자료조사 후 잠재적인 유해성의 모든 징후도 보고한다. 이러한 잠재적인 유해 효과의 실상에 대한 평가는 수많은 비교로 인한 다중성 문제를 고려한다. 이상반응 발생률과 노출 및 추적기간과의 가능한 관련성을 밝히기 위해 생존분석을 적절히 사용한다. 확인된 유해효과와 관련된 위험도는 적절한 위험/편익 평가가 가능하도록 정량화 한다

8. 참고 : 임상시험 통계용어 해설

계획서 순응 임상시험대상자군, Per Protocol Set

과학적 모델을 이용하여 분석하였을 때 이들 자료로써 충분히 치료효과를 나타낼 수 있도록 임상시험 계획서에 순응하여 연구를 종료한 연구대상 집단에서 얻어진 자료를 말한다. 여기서 순응도는 치료의 노출에 대한 고려, 측정치의 활용가능성, 그리고 주요한 임상시험 계획서 위반사항이 없다는 것을 포함한다.

권장용어와 유사용어, Preferred and Included Terms

유사용어는 연구자가 약물 부작용을 기술하는데 사용하는 각종 다른 용어를 코드화하는데 가장 하위 수준의 용어이고, 권장용어는 전통적으로 부작용 발생의 빈도를 보고하는데 대표적으로 사용하는 용어를 말한다(MedDRA와 같은 계층적인 의학용어 사전에서 분류하여 기록됨). 예를 들면, 연구자가 기술한 경우에 유사용어로 'point pain'으로 코드화 될 수 있으나, 권장용어로는 'Arthralgia'로 보고된다.

내용 타당도, Content Validity

어떤 변수가 측정하고자 하는 내용을 얼마나 정확히 측정하였는 지의 정도를 나타내는 지표이다.

눈가림 검증(눈가림 조사), Blind Review

계획된 분석을 종료할 목적으로 임상시험의 종료시점과 눈가림 해제사이 기간 동안에 연구자료를 검토·평가하는 것을 말한다.

다기관 임상시험, Multicentre Trial

하나 이상의 기관에서 단일 계획서에 따라 수행된 임상시험을 말하며, 다기관에서 행해지므로 한 명 이상의 연구자에 의해 수행된다.

대리변수, Surrogate Variable

임상 효과의 직접적인 측정이 쉽지 않거나 불가능한 상황에서 간접적으로 효과를 측정해주는 변수를 말한다.

독립적 자료 모니터링 위원회(IDMC), Independent Data Monitoring Committee

독립적 자료 모니터링 위원회는 임상시험의 진행, 자료의 안전성, 임상적인 유효성 평가변수를 정기적으로 평가하기 위해 의뢰자에 의해 설립될 수 있고, 이 위원회에서는 임상시험의 지속, 변경, 중단 여부에 대해 의뢰자에게 조언한다.

동등성 평가시험, Equivalence Trial

두 가지 혹은 그 이상의 치료 반응이 임상적으로 중요하지 않은 양만큼 차이가 남을 입증하는 것을 주 목적으로 하는 임상시험을 말한다. 보통 이것은 진정한 치료차이가 임상적으로 받아들일 수 있는 차이의 상한과 하한의 범위내에 있다는 것을 보임으로써 입증할 수 있다.

메타분석, Meta-Analysis

동일한 주제에 대해 수행된 두 가지 이상의 임상시험으로부터 정량적인 증거를 평가할 목적으로 행하는 분석을 말한다. 일반적으로 다른 여러 임상시험의 결과로 제시된 요약 통계량을 통계적으로 결합하는 과정을 가장 흔히 사용하지만, 때로는 원 자료를 결합하여 분석하는 것을 의미하기도 한다.

무작위 배정된 모든 임상시험대상자 군, Full Analysis Set

‘배정된 대로 분석’ 원칙에 의해 제공되는 개념과 가능한 한 근접하는 이상적인 연구대상 집단을 말한다. 이것은 모든 무작위 배정된 대상자로부터 최소한의 제외사유가 정당한 연구대상자를 제거한 임상시험대상자군이 된다.

‘배정된 대로 분석’ 원칙, Intention-To-Treat Principle

무작위 배정된 모든 연구대상자는 처음 배정된 치료군에 속한 것으로 하여 주 분석을 시행하여야 한다는 원칙이다. 즉, 치료군에 배정된 연구대상자는 계획된 치료에 순응하는지에 상관없이 배정된 군의 구성원으로 추적 관찰되고, 평가되고, 분석되어야 한다는 것이다.

베이저안 접근법, Bayesian Approaches

미지의 모수(예, 치료효과)에 관한 실제 관측이전의 지식을 사전분포라는 확률

분포로 나타내고 여기에 관찰된 자료를 결합하여 사후분포를 제공해주는 자료 분석법을 말한다. 여기서 구한 사후분포는 미지의 모수에 관한 모든 통계적 추론을 위한 근거로 사용된다.

비뚤림(통계적 & 실행적), Bias(Statistical & Operational)

치료효과를 추정하기 위한 임상시험의 계획, 수행 및 결과해석 과정에서 치료 효과의 추정치를 참 값과 벗어나게 만드는 요소들의 계통적 경향성을 말한다. 연구수행상의 이탈로부터 기인한 비뚤림을 ‘실행적 비뚤림’ 이라 하며, 위에서 열거한 다른 원인인 분석과 결과해석 단계에서 기인한 비뚤림을 ‘통계적 비뚤림’ 이라 한다.

예를 들어, 치료군 배정시 예후가 나쁜 임상시험대상자들이 계획적으로 한가지 치료법을 지정받는 것은 조작적 비뚤림이고, 계획서 위반이나 임상시험대상자의 결과에 근거하여 분석에서 제외시킴으로써 정확한 치료효과를 평가하는데 영향을 미치는 것은 통계적 비뚤림이다.

비열등성 평가시험, Non-Inferiority Trial

새로운 치료약의 결과가 비교하는 약(기존 치료약이나 위약)보다 임상적으로 열등하지 않음을 입증하는 것을 주목적으로 하는 임상시험을 말한다.

상호작용(정성적 & 정량적), Interaction(Qualitative & Quantitative)

상호작용은 치료효과의 차이가 연구대상이 되는 약물이 아닌 다른 요인에 의해 영향을 받는 상황을 말한다. 예를 들어 치료군과 대조군간 치료효과의 차이가 각 센터에 따라 변화되는 상황을 말한다. 정량적 상호작용은 치료효과 차이의 크기가 요인의 수준에 따라 다르게 나타나는 상황을 말하며, 정성적 상호작용은 요인의 적어도 한 수준에서 치료효과 차이의 방향이 다르게 나타나는 것을 말한다.

안전성과 내약성, Safety & Tolerability

안전성은 일반적으로 임상검사, 활력징후(vital signs), 임상적인 이상반응(질병, 증상 및 증후) 그리고 다른 특수 안전성 검사(예, 심전도 검사, 안과검사 등)에 의해 임상시험에서 평가되는 연구대상자의 의학적 위험도와 관련이 있다.

내약성은 임상적으로 명백한 이상반응을 연구대상자가 참아낼 수 있는 정도를 나타낸다.

우월성 평가시험, Superiority Trial

새로운 치료약의 결과가 비교대조 약(기존 치료약이나 위약)보다 우월함을 입증하는 것을 주목적으로 하는 임상시험을 말한다.

이중위약, Double-Dummy

두 치료제가 동일하지 않을 때 눈가림법을 유지하기 위하여 사용하는 방법이다. 치료제 A에 대한 위약과 치료제 B에 대한 위약을 준비하여, 각 연구대상자에게 A 치료제와 B위약 또는 A위약과 B치료제 두가지 경우의 조합 중 한가지를 배정하게 된다.

일반화, Generalisability, Generalisation

임상시험의 결과를 시험에 참여한 연구대상으로부터 더 넓은 범위의 환자집단과 임상상황에까지 적용시킬 수 있는 정도를 말한다.

종합평가변수, Global Assessment Variable

환자의 상태나 그 변화에 대한 객관적인 변수와 연구자의 전반적인 의견을 통합하는 단일 변수를 말하며, 일반적으로 순서화된 범주형 척도로 나타낸다.

중간분석, Interim Analysis

임상시험을 공식적으로 종료하기 전 어느 시점에서나 치료군간에 유효성이나 안전성을 비교·평가하기 위하여 시행하는 분석을 말한다.

중도탈락, Dropout

임상시험의 연구계획서에서 요구된 마지막 방문까지 계속할 수 없는 이유를 가져 관찰이 중단된 연구대상을 말한다.

치료효과, Treatment Effect

임상시험의 치료로 인해 나타난 효과를 말한다. 대부분의 임상시험에서 연구

대상 치료효과는 2개 이상의 치료군에 대한 비교에 있다.

치료 중 발생한 이상반응, Treatment Emergent

치료전에는 없었던 반응이 치료 도중 나타났거나, 어떤 반응이 치료전의 상태보다 더 악화된 경우를 말한다.

통계분석계획, Statistical Analysis Plan

통계분석계획은 임상시험 계획서에서 언급된 분석의 주요한 특성을 보다 기술적이고 상세하게 포함하고 주 변수와 이차변수의 통계적 분석 수행을 위한 보다 자세한 처리절차를 포함하는 내부문서를 말한다.

평가자간 신뢰도, Inter-Rater Reliability

서로 다른 평가자가 각기 다른 상황에서 평가하였을 때 그 결과가 동일하게 나타나는 정도를 말한다.

평가자내 신뢰도, Intra-Rater Reliability

같은 평가자가 다른 상황에서 평가하였을 때 결과가 동일하게 나타나는 정도를 말한다.

확률 통계적 방법, Frequentist Methods

동일한 실험 조건이라는 가정하에서 얻어지는 결과의 빈도를 이용하여 유의성 검정과 신뢰구간을 구하는 등의 통계적 방법을 말한다.

제·개정 이력

연번	제·개정번호	승인일자	주요내용
1	B1-2015-2-009	2015. 9.	제정

임상시험의 통계 원칙

발 행 일 2015년 9월

발 행 인 손 여 원

편 집 위 원 장 이 선 희

편 집 위 원 최영주, 김호정, 손경훈, 김희성, 이윤숙, 고용석, 김영주,
박봉서, 도희정, 정명아, 구민지, 송주경, 조철호, 김은영

발 행 처 식품의약품안전평가원 의약품심사부 의약품심사조정과
