

임상시험계획서 작성 관련 통계 FAQ*

국립독성과학원 응용통계과

* 본 FAQ는 응용통계과에서 '08년도 3분기까지 수행한 통계기술지원 결과를 토대로 다빈도 검토내용을 질의/답변 형식으로 작성함

Q. 임상시험계획에 포함되어야 할 통계적 고려사항은 무엇인가요?

A. 통계적 측면에서 임상시험계획서에 포함되어야 할 내용에는 연구목적에 부합하는 연구 설계방법, 유효성 변수의 정의, 통계적 가설, 피험자 수 산출방법, 통계적 분석방법 등이 있다. 비교임상인 경우 대조군의 성격, 무작위 배정방법 등을 고려해야 하며 적절한 유효성 평가변수를 정의하고 이를 토대로 타당한 방법에 의하여 피험자 수를 제시해야 한다. 변수의 성격과 연구가설에 적합한 분석방법을 제시하고 이 때 필요한 가정들이 만족되는 지를 검정하는 절차를 밝아야 한다.

Q. 비열등성 시험이 필요한 경우는 언제인가요?

A. 비열등성 시험은 임상적으로 허용할 수 있는 한계의 범위에서 시험군의 치료효과가 대조군의 효과보다 나쁘지 않음을 보이는 임상시험이다. 비열등성 시험은 대조군에서 위약을 사용할 수 없어 표준시험약을 사용하는 임상시험에 적용할 수 있다. 비열등성 시험에서 대조군보다 나쁘지 않음을 허용할 수 있는 수준을 비열등성 한계라고 하며 이는 통계적 측면, 임상적 측면에 근거해서 신중하게 결정되어야 하며 그 근거가 임상시험계획서에 기술되어야 한다.

Q. ITT분석과 PP분석은 무엇인가요?

A. 피험자는 우발적으로 또는 고의적으로 연구계획서를 따르지 않을 수 있고 임상 의사는 환자의 상황에 따라 원래 배정된 치료와 다르게 치료하는 경우가 있을 수 있다. 이를 처리하는 방법으로 ITT 분석(intention to treat analysis)은 모든 무작위 배정된 환자들을 배정된 대로 유지한 채 자료를 분석하는 것이다. 이것은 무작위 배정을 유지하기 때문에 그룹 간 균형을 유지하게 한다. ITT 분석원칙은 임상시험계획서에 명시되어야 한다. 반면 PP분석(per protocol analysis)은 ITT에 포함된 피험자 중에서 임상시험계획서에 따라 임상시험을 완료한 피험자로부터 얻어진 자료를 분석에 포함하는 방법이다.

Q. 일차 유효성 평가변수와 연구가설에서 주의할 사항은 무엇인가요?

A. 일차 유효성 평가변수(primary variable, primary endpoint)는 임상시험의 유효성을 평가하는 척도로서 임상시험의 주요 목적과 연관성이 높아야 하며 신뢰성과 타당성 있게 선택되어야 한다. 임상시험에서 증명하고자 하는 연구가설은 유효성 평가변수에 근거하여 통계적 가설의 형태로 명확하게 정의되어야 하며 이를 근거로 임상시험에 필요한 피험자 수를 산출해야 한다. 비교 임상시험인 경우 대조군이 위약을 사용하는지, 표준약을 사용하는지를 고려하여 연구가설을 세워야 한다. 특히 가설의 귀무값(귀무가설에서 제시되는 값)은 유효성을 입증하는 기준이 되는 값이므로 매우 신중하게, 전문가적인 입장에서 결정되어야 한다.

Q. 피험자 수 산출시 고려사항은 무엇인가요?

A. 가설검정을 기초로 한 피험자 수 산출은 임상시험 설계방법에 따라 유의수준, 제2종 오류의 확률 하에서 계산한다. 유효성을 입증하는 가설의 귀무값, 제2종 오류를 허용할 수 있는 시험약 효과에 대한 사전 정보가 있어야 한다. 흔히 귀무값과 제2종 오류를 허용할 수 있는 시험약 효과와의 차이를 “효과의 차이”라고 한다. 신뢰구간을 토대로 한 피험자 수 산출은 신뢰수준과 추정오차에 대한 정보, 신뢰구간의 오차의 한계가 주어져야 한다. 피험자 수 산출시, 선행 연구에서 유의한 결과를 얻은 피험자 수와 유사한 수준에서 임상시험을 계획하는 것은 바람직하지 않다. 이것은 무작위 표본추출로 인한 ‘행운’일 가능성이 있으므로 선행 연구결과를 토대로 적절한 방법에 근거하여 피험자 수를 산출해야 한다. 발표된 사전 정보가 없다면 예비연구를 수행할 수 있다. 예비적인 성격의 pilot 연구로서 연구의 실행 가능성을 평가하거나 정보를 수집하는 연구에서는 피험자 수 산출이 필요하지 않으나 이러한 연구의 목적을 명확히 기술해야 한다. 또한 이용 가능한 환자 규모를 고려한 피험자 수 산출은 적절하지 않고, 임상적으로 유의한 효과를 입증하기에 충분한 검정력을 갖지 못할 정도로 환자의 규모가 작다면 연구 기간을 확장하거나 다기관 공동연구 수행을 고려해야 한다.

Q. 무작위배정이 필요한 이유는 무엇인가요?

A. 무작위배정은 임상시험 과정에서 발생할 수 있는 비뚤림(bias)을 줄이기 위해 확률의 원리에 따라 피험자를 각 치료군에 배정하는 과정이다. 무작위 배정을 통해 피험자 특성의 차이가 유효성 효과와 교락되는(confounding) 것을 방지할 수 있다. 또한 무작위 배정을 통하여 랜덤표본(random sample)이 추출되도록 함으로써 통계적 가설검정의 기본 가정을 충족할 수 있다. 임상시험계획서에는 무작위배정을 어떠한 방식으로 수행할 것인지를 명시해야 한다.

Q. 층화 무작위 배정은 언제 필요한가요?

A. 일차 유효성 평가 변수에 영향을 미칠 것으로 예상되는 예후변수(prognostic variable)에 대해서는 층화 무작위 배정을 하는 것이 바람직하다. 층화 무작위 배정은 유효성 평가변수에 영향을 미칠 것으로 예상되는 변수를 층(stratum)으로 나누고 각 층 내에서 무작위 배정표를 이용하여 피험자를 무작위 배정하는 것이다. 몇 개의 층으로 구성할 것인지를 임상시험계획서에 명시해야 한다.

Q. 비뚤림이 발생하는 이유는 무엇인가요?

A. 비뚤림은 일차 유효성 평가 변수의 측정값이 참값에서 멀어지는 현상이다. 비뚤림은 피험자의 잠재의식에 따라 일어날 수 있고 어떤 치료를 받고 있는지에 대한 환자의 인지 또는 치료에 대한 믿음에 따라 측정값에 영향을 미칠 수 있다. 따라서 피험자에 대한 맹검이 깨지게 되면 비뚤림 현상은 크게 나타날 수 있다.

Q. 맹검 시험은 무엇인가요?

A. 맹검 시험은 임상시험에 관여하는 사람 또는 부서 등이 배정된 치료법에 대해 알지 못하도록 하는 절차이다. 단일 맹검 시험은 피험자가 치료법에 대하여 알지 못하나 연구자가 치료법에 대하여 알고 있는 시험이다. 피험자가 치료법에 대한 정보가 없어 위약효과가 없을 것이라고 기대될지 모르나 연구자의 피험자를 대하는 태도라든지 비양심적인 힌트 제공 등을 통해 유효성 평가 변수에 영향을 미칠 수 있다. 이중 맹검 시험은 피험자와 연구자 모두 치료법에 대하여 알지 못하는 시험으로 비뚤림을 방지할 수 가장 좋은 시험방법이다.

Q. 위약효과는 무엇인가요?

A. 이중 맹검 임상시험에서 피험자는 치료의 종류에 대해 식별할 수는 없지만 특정 치료를 받는 것으로 얹으로써 자신의 상태가 호전되었다고 느끼는 심리적인 반응을 위약효과(placebo effect)라고 한다.

Q. 유의확률 보정이 필요한 이유는 무엇인가요?

A. 제1종 오류는 실제로 유의한 차이가 없음에도 불구하고 유의한 차이를 얻게 될 오류이며, 유의확률(p-value)은 제1종 오류를 범할 최대 확률이다. 유의성 검정에서는 오류를 범할 가능성을 유의수준(일반적으로 5%) 이하로 제한한다. 그러나 여러 번의 유의성 검정을 수행하게 되면 오류의 가능성은 증가한다. 예를 들어 2회의 유의성 검정에서 제1종 오류의 가능성은 $1-(1-0.05)^2=0.0975$, 3회의 유의성 검정에서 오류의 가능성은 $1-(1-0.05)^3=0.143$ 로 증가한다. 따라서 다중비교, 중간분석과 같이 독립적으로 수행해야 하는 유의성 검정의 횟수가 증가하면 유의확률의 보정을 통해 제1종 오류의 발생 가능성을 조정해 주어야 한다.

Q. 다중검정 문제는 언제 발생하나요?

A. 유의성 검정을 여러 번 실시하는 다중검정에서는 제1종 오류의 확률이 증가하게 된다. 이를 다중검정의 문제라 하며 이러한 문제의 발생 원인은 첫째, 하나의 임상시험에서 결과변수의 수가 두개 이상이어서 여러 번 검정을 수행해야 하는 경우이다. 이 경우 Bonferroni 방법 등을 사용하여 유의성 검정을 실시할 수 있다. 둘째, 시간에 따라 반복 측정된 자료분석에서 각각의 시간대에서 개별적으로 검정을 하는 것은 제1종 오류의 가능성을 증가시킨다. 반복 측정자료 분산분석법 등을 적용하여 발생하는 문제를 해결할 수 있다. 셋째, 임상시험 군이 세군 이상일 때 군간 비교하는 경우이다. Bonferroni 방법 등을 사용할 수 있으나 검정력이 떨어질 수 있음에 주의해야 한다. 다중검정을 하게 되는 경우 임상시험계획에서 사용하고자 하는 방법을 명시하고 이를 따르는 것이 바람직하다.

Q. 유의확률 보정방법에는 무엇이 있나요?

A. 가장 쉽게 적용할 수 있는 유의확률 보정방법은 Bonferroni 방법이다. 유의수준이 α 인 경우, Bonferroni 방법은 k 개의 유의성 검정 중 각 검정에서의 유의확률을 α/k 와 비교하는 방법이며, 모든 k 개의 유의성 검정에서 제1종 오류의 가능성은 근사적으로 α 가 된다. 이 밖에도 오류의 가능성을 증가시키는 원인에 따라 다양한 접근방법을 적용할 수 있다.

Q. 교란변수와 교호작용은 어떻게 다른가요?

A. 교란변수(confounding factor)는 결과변수와 인과적으로 관련성을 가지면서 설명변수와는 비인과적 연관성을 가지는 변수로, 일반적으로 관찰연구에서 나타나는 문제이다. 임상시험은 교란변수를 없애거나 보정할 수 있도록 설계되어야 하며, 교란변수가 관심요인과 결과변수의 연관성 감지에 영향을 주지 않도록 해야 한다. 교호작용(interaction effect)은 두 변수 간 연관성의 특성이 또 다른 변수에 따라 달라질 때 발생한다.

Q. 다기관 임상시험의 경우 기관 간 효과를 통계분석에 포함해야 하나요?

A. 다기관 임상시험은 하나의 임상시험계획서에 따라 두개 기관 이상의 기관에서 수행되는 임상시험을 말한다. 기관을 층화 무작위에 사용하는 경우 기관 변수를 공변량으로 한 공분산 분석을 실시해야 한다.

Q. 공분산 분석은 언제 하나요?

A. 공변량 변수를 포함하는 공분산 분석은 임상시험계획서에 명시된 경우에만 수행하는 것이 바람직하다. 층화 무작위를 실시한 경우 유효성 평가변수 분석에서 층화에 사용된 변수를 보정하는 공분산분석을 실시해야 한다. 유효성 평가변수와 관련이 있는 것으로 알려진 예후 변수를 공변량으로 사용하기 위해서는 임상시험계획서에 이를 기술해야 한다. 그러나 기초조사(Baseline)에서 불균형으로 관측된 변수를 공변량으로 포함하는 것은 적절하지 않다.

Q. 자료맹검과 SAP(통계분석계획서)는 무엇인가요?

A. 자료맹검은 임상시험을 통해 얻은 자료로부터 임상시험의 결과를 확인하지 못하도록 피험자, 연구자에게 자료를 공개하지 않는 것으로, 임상시험이 모두 끝나서 자료 분석을 시작할 때까지 맹검을 유지하는 것이 바람직하다. SAP(통계분석계획서)는 임상시험계획에 따른 자료분석에 대한 통계적 방법을 임상시험계획서와는 별도로 작성한 계획서이다. 임상시험계획서와 달리 통계분석계획서에서의 분석 방법은 임상시험 수행 중에 수정이 가능하나 자료의 맹검이 깨지는 순간부터 수정할 수 없다.