

임상시험에서의 통계 활용

안 홍 엽

동국대학교-서울 통계학과

Usage of Statistics in Clinical Trials

Hong Yup Ahn, Ph.D.

Department Statistics, Dongguk University-Seoul, Seoul, Korea

The purpose of clinical trials is to find evidences for the effects of experimental new drugs or treatments on human. For the successful clinical trials, it is not sufficient to use statistics only for the analyses of collected data, but it is necessary to extend the usage of statistics in various ways. At the beginning of the study, one needs to use statistics for systematically and concretely planning the study. For this, we discussed the usage of statistics in defining the effect, determining the sample size, statistical analyses, and so on. (Korean J Hosp Palliat Care 2010;13:1-6)

Key Words: Clinical trials as topic/statistics & numerical data, Sample size, Data interpretation, Statistical

서 론

Statistics는 1963년 인구통계학자 John Graunt의 Natural and Political Observations upon the Bills of Mortality에서 처음 소개되었다(1). 이 당시의 통계는 현대적인 자료 분석을 위한 통계모형 또는 분석방법을 제시하는 수준은 아니었지만 정책을 결정하는데 현 상황을 파악하는 것이 필요하며 이를 위해 자료를 수집 정리하는 것이 요구된다는 사실을 인지했다는 점에서 통계학사에 큰 의미가 있다고 하겠다. 초보적인 자료 수집 및 정리에서 시작한 통계학은 이론적인 발전을 바탕으로 현재와 같은 중요한 학문으로 인정받게 되었다. 17세기부터 시작한 Blaise Pascal과 Pierre de Fermat의 확률론은 통계학에 필수적인 수학적 이론을 제공하였으며 1974년 Carl Friedrich Gauss의 최소자승법은 다양한 통계 모형을 개발하는데 활용되었다(2).

접수일: 2010년 2월 10일, 수정일: 2010년 2월 10일

승인일: 2010년 2월 10일

교신저자: 안홍엽

Tel: 02-2260-8585, Fax: 02-2267-0998

E-mail: ahn@dongguk.edu

통계를 이용한 자료 분석 및 분석 결과의 해석은 다양한 연구에서 연구 가설을 증명하는데 사용되고 있다. 진단 검사 방법들의 효용성을 과거 선행 연구들의 결과들로부터 비교하는 메타분석, 암환자의 생존 시간에 영향을 주는 요인 탐색을 위한 생존분석, metabolic syndrome에 관한 연구, 지방간과 acute ischemic and hemorrhagic stroke 간의 상관관계 등의 연구에서 통계 모형을 이용한 통계 분석은 연구 가설을 증명하는데 필수적이면서 동시에 효과적으로 과학적 증거를 제시하기 위해 사용되고 있다.

다양한 연구에서 통계의 수요는 매우 높아지면서 통계에 대한 인식의 변화가 보이고 있다. 주어진 자료에 대한 통계 분석만으로 통계의 활용을 제한하기 보다는 연구 전반에 걸쳐 통계 분석을 염두해 두고 더욱 치밀한 연구계획을 세우는 것이 요구되고 있다. 이는 특히 임상시험에서 강조되고 있다. 많은 임상시험의 경우 연구 계획 단계에서부터 통계의 활용을 고려하는 것이 전체 임상시험을 효율적으로 계획할 수 있음이 확인되었다(3).

Bearman(4)은 임상시험에서의 연구 계획서는 다음과 같이 활용된다고 정리하였다.

1. 연구 실시 전, 연구 계획서는 연구가 지향하는 바와 최종 목적지까지 도달할 수 있는 방법을 제시한다.
2. 연구 수행 중, 연구 계획서는 연구의 진행 정도와 앞으로 진행해야 할 부분에 대한 판단 척도로서 활용된다.
3. 연구 종료 후에는 연구 계획서를 바탕으로 연구 진행 과정에 대한 전반적인 검토를 실시할 수 있다.

본 논문에서는 임상 시험의 연구 계획서 작성시 통계 관련 사항으로 고려되는 다음 4가지 사항을 중심으로 알아보도록 한다.

1. 연구 가설을 확인하기 위해 관측할 변수들의 결정
2. 자료 측정에 관한 시험설계
3. 적절한 표본크기 산출
4. 적절한 통계 모형과 이를 바탕으로 한 유효성 판정 기준

본 론

1. 관측 변수

특정 약물이 환자에게 도움이 되는지를 확인하고자 하는 임상시험을 생각해 보자. 임상시험을 통해 적절한 연구 대상자로부터 여러 가지 관측값을 측정할 수 있다. 다양한 관측값은 실험의 결과를 계량화 시키는데 많은 도움을 줄 수 있지만 너무 많은 변수에 대해 관측을 계획한다면 필요 이상으로 시간과 노력을 낭비하게 될 수 있다.

관측되는 변수는 그 활용 목적에 따라서 약물의 효과를 보여주는 결과값에 해당되는 효과평가변수와 효과평가변수에 영향을 줄 수 있는 관심변수와 교란변수로 구분된다. 여기서 관심변수는 효과평가변수에 영향을 준다는 점에서 교란변수와 유사하다. 하지만 효과평가변수에 대한 관심변수의 영향이 직접적인 연구 가설에 관련되는 것에 반해 교란변수는 관심변수의 영향력을 보정해 주는 보조적인 역할에 그친다는 점에서 차이가 있다. 효과평가변수의 결정은 매우 중요한 사항으로 충분한 문헌 고찰과 공동연구자들의 다양한 의견교환을 통해 결정되어야 한다. 교란변수는 대상자의 성별, 연령, 나이, 가족상태 등의 인구통계학적 변수들과 키, 체중, 가족력, 혈액검사 자료 등과 같은 효과평가변수와 간접적인 관련성이 있을 것으로 예상되는 항목으로 이루어져 있다. 연구자는 이러한 점을 유의하여 주어진 자원에서 가장 효율적인 연구가 진행될 수 있도록 관측 변

수를 정해야 한다.

많은 경우 연구자는 임상시험을 통해 시험 대상 약물이 통계적으로 유의한 효과를 보여주기를 바란다. 때때로 유의한 효과를 보여주지 못한 임상시험도 발생하는데 이러한 결과 자체는 연구라는 측면에서는 나름대로 의미를 갖는다.

하지만 현실적으로 유의한 연구 결과를 추구한다는 점을 생각해 볼 때, 효과평가변수의 잘못된 선택으로 효과가 없는 것으로 단정 지을 가능성도 배제해서는 안 된다. 적절한 효과평가변수 결정은 ‘유의한 효과 차이’를 어떻게 정의하는가에 달려 있다.

유의한 효과 차이는 다양하게 정의될 수 있다. 시험군과 대조군의 효과평가변수의 평균값에 유의한 차이가 있는 것을 흔히 유의한 효과 차이가 있는 것으로 판정한다. 하지만 평균값의 비교만으로 부족한 경우가 있을 수 있다. 예를 들어 감기를 앓고 있는 환자에게 약물 A를 처방한 경우 ‘평균’적으로 10일이면 완치가 된다고 하자. 여기서 ‘평균’이 그 약을 처방 받은 환자가 10일 정도면 반드시 완치된다는 것을 보장하는 것은 아니다. 환자에 따라서는 한 달 내내 고생할 수도 있는 것이다. 다만 여러 환자들에게 처방한 경우 그들의 완치 기간의 평균값이 10일 정도라는 것을 의미하는 것이다. 그런데 약을 처방 받지 않고 집에서 휴식을 취하는 경우 완치 기간 역시 평균적으로 10일이라고 한다면 평균 완치 기간의 차이가 없으므로 약물 A는 감기에 도움이 되지 않는다고 판정할 것이다. 평균값에 따른 효과 차이를 정의한다면 당연한 결과이다.

효과차이를 효과평가변수의 평균 비교로 정의 내리는 것만은 아니다. 앞의 감기약에 관한 예에서 이번에는 표준편차(또는 분산)를 생각해 보자. 임상시험 결과 약물 A를 처방 받은 환자의 완치 기간의 평균 일수는 10일이었지만 표준편차가 1일이었다고 하자. 그리고 집에서 단순히 휴식을 취한 환자의 경우 평균 완치일수는 10일로 동일하지만 표준편차는 5일이었다고 하자. 비록 평균 완치일수에서는 차이가 없었지만 표준편차는 차이가 있는 것이 분명하다. 이 경우 약물 A는 유의한 효과차이가 있다고 판정할 수 있다.

임상시험에서 특정 약물의 효과를 확인하였다 해서 환자 개인에게 반드시 효과가 나타난다고 보장할 수는 없다. 단지 임상적인 결과가 더 좋게 나타날 ‘가능성’이 높다는 것을 의미하는 것이다. 통계에서는 그 가능성의 척도로 확률을 사용하고 확률과 관련성이 매우 높은 통계량으로서 평균과 표준편차를 이용하여 효과의 비교

를 실시한다. 앞의 감기의 경우 정규성 가정 하에서 약물 A의 완치일수가 10일 이하일 확률은 50%로 집에서 휴식을 취하는 경우와 동일하다. 이를 Figure 1A를 통해 설명하면, 실선은 약물 A를 사용한 경우의 완치일의 확률분포를 나타내고 있으며 회색부분의 면적은 완치일이 10일 이하일 확률을 의미한다. 마찬가지로 점선은 휴식만을 취한 경우의 완치일의 분포를 나타낸 것으로 빗금친 영역의 면적은 휴식만을 취한 경우 완치일이 10일 이하일 확률을 의미한다. 모두 정규분포를 가정하고 평균 이하의 확률이 50%이므로 회색영역과 빗금 친 영역의 면적은 50%로 동일하다. 하지만 완치일수가 평균값인 10일보다 단지 1일 더 많은 11일 이하일 확률은 약물 A를 사용하는 경우가 집에서 휴식을 취하는 경우보다 더 높다(Figure 1B). 그리고 그 확률의 차이는 11일보다 12일 이하일 때 더 커진다. 즉 평균을 지나서 비교 기준일이 더 늘어날수록 약물 A의 효과는 더욱 긍정적인 효과를 나타낼 가능성이 커지는 것을 알 수 있다. 따라서 이러한 확률분포의 성질을 이해하고 있다면 약물 A가 휴식과 동일한 평균 완치일을 갖기 때문에 효과가 없다고 단정하는 것은 적절치 않다는 것을 이해할 수 있다.

효과평가변수가 만일 계량적으로 측정되는 경우가 아닌 범주형인 경우는 효과의 차이는 다르게 정의된다. 특정 질환의 진단을 위한 새로운 검사법이 제안된 경우 기존 검사법보다 더 좋은 성능을 지니고 있는가를 확인하는 경우를 생각해 보자. 이러한 경우 효과평가 변수는 검사 결과가 양성/음성과 같은 범주형 자료이다. 범주형 자료이기 때문에 앞의 감기의 예처럼 자료값의 평

균 또는 표준편차를 이용한 비교는 불가능하다. 따라서 범주형 자료는 측정 결과를 양성 반응 수 또는 비율과 같이 먼저 계량화 시킨 후 통계 분석을 실시하게 된다. 많은 경우 효과의 차이를 비교하기 위해 오즈비, 민감도, 특이도 등과 같은 다양한 통계량을 연구 목적에 맞게 적절히 정한 후 유의한 효과 차이를 정의한다.

2. 시험설계

시험의 종류는 크게 관측연구(observational study)와 실험연구(experimental study)로 나눌 수 있다. 두 시험의 형태는 연구자의 개입(intervention)이 존재하지 않은 경우 관측연구, 존재하는 경우 실험연구로 구분된다. 예를 들면 Kaku와 Lowenstein(5)는 15세에서 44세 성인들을 대상으로 약물 남용(recreational drug abuse)과 stroke 간의 관계를 조사하였다. 이 경우 윤리적인 문제 때문에 연구 대상자가 약물 남용을 할지 아니면 하지 않을지를 연구자가 지정할 수 없다. 따라서 이 연구는 관측연구가 된다. 반면에 Steering Committee of the Physicians' Health Study Research Group(6) 아스피린이 심혈관계 질환 발생을 줄여주는 효과를 확인하는 임상시험을 실시하였다. 이 경우에는 22,000명의 시험 대상자들에게 연구자가 무작위 배정을 통해 아스피린 또는 위약을 배정하여 시험이 실시되었다.

임상시험(clinical trial)은 실험연구의 한 형태로, 인간을 대상으로 특정 약물 또는 치료법의 효과를 밝히는 연구 목적을 갖는 시험이다. 임상시험은 시험 대상자를 시험군과 대조군으로 나누어 시험을 실시하는 RCT (randomized clinical trial), 동일 대상자에 대해 시험군과 대조군

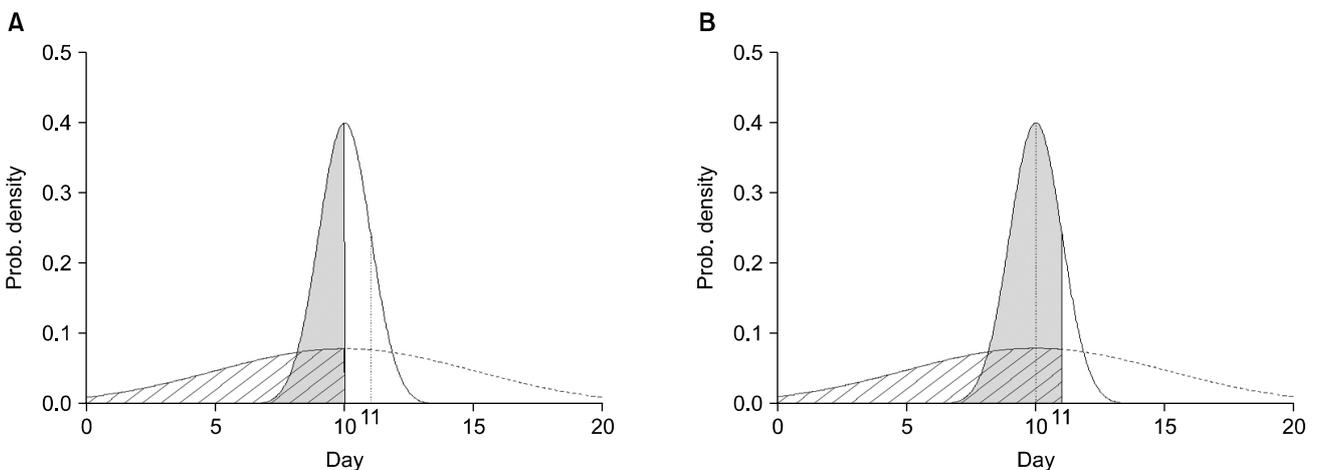


Figure 1. Normal distribution. Solid line is for normal distribution with mean 10 and sd 1. Dashed line is for normal distribution with mean 10 and sd 5. the shaded and dashed areas of (A) and (B) are corresponding to the probabilities that day is less than 10.

에 순차적으로 배정하는 crossover study, 그리고 시험군만으로 시험을 실시하고 대조군은 이미 이루어진 선행 연구의 결과로 같음하는 경우로 나누어 생각해 볼 수 있다. 이 중에서 임상시험에 무작위 배정을 통해 시험군과 대조군의 효과 비교를 통해 연구 가설을 확인하는 RCT는 인과관계를 밝히는 연구 시험 중에서 가장 좋은 그야말로 여러 시험방법들 중의 정수라 할 수 있다(7).

3. 표본크기

특정 약물이 환자에게 도움이 되는지를 확인하고자 하는 임상시험을 생각해 보자. 우리는 우선 누구를 대상으로 임상시험을 실시할지를 정해야 한다. 간단한 예로 금연패치의 효과를 확인하는 임상 시험에서 시험 대상자의 대부분이 금연패치 제조사의 직원이거나, 또는 비흡연자라면 그 임상 시험은 적절히 실시된 것으로 보기 어려울 것이다. 적어도 시험 대상자는 임상시험의 목적에 부합되고, 시험에 편향된 관측값을 제공할 가능성이 적은 대상으로 선정되어야 할 것이다. 대상자 선정에 관한 기준은 연구 계획서 상에 inclusion/exclusion criterion을 제시하여 연구가 종료 될 때까지 지속적으로 연구 대상자 선정에 참고하도록 한다.

적절한 시험 대상자가 결정되면 이들 중 몇 명을 대상으로 임상시험을 실시할지를 정해야 한다. 임상시험을 계획하는데 있어 중요한 단계 중의 하나인 표본크기를 결정하는 것은 경제적으로나 윤리적으로 큰 의미를 갖는다. 사람을 대상으로 하는 시험에 불필요하게 많은 사람들을 참여시켜 연구 기간이 늘리는 경제적 손실은 당연히 피해야 한다. 시험을 통해 검정되지 않은 약물 또는 치료 방법의 효과를 검정하는 것이 임상시험의 목적이라는 점을 상기해 볼 때 불필요하게 많은 시험 대상자에게 부작용 등의 위험에 노출시키는 것은 윤리적으로 옳지 않다. 따라서 시험에서 요구 되는 최소 표본크기를 산정하여 위험 정도를 최소화 시키는 것이 필요하다. 물론 최소화된 '희생'을 통해 인류에게 도움이 되는 연구 결과를 얻는 것이 임상시험의 목적이라는 점을 감안한다면 이러한 희생은 필수불가결하다고 하겠다.

반대로 시험에서 요구되는 표본크기 보다 부족한 시험 대상자를 확보하여 시험을 실시하는 경우 단순히 표본크기가 충분하지 않기 때문에 유의한 효과가 증명되지 않을 가능성이 있다. 이 경우 시험에 참여한 사람들을 위험에 노출시켰음에도 불구하고 유효한 연구 결과를 도출하는데 실패함으로써 연구는 이득 없는 자원 낭비의 예로 전락하게 된다.

일반적으로 표본크기를 산정하는데 다음과 같은 사항을 고려하게 된다(8).

1. 시험의 형태
2. 효과평가변수와 효과크기 판정 방법
3. 효과평가변수의 최대 허용 오차
4. 효과평가변수의 산포도(variability)
5. 유의수준 및 검정력
6. 적절한 통계방법에 의한 최소 표본 크기 산정

시험의 형태로는 대체로 시험 대상자를 각각 시험군 또는 대조군에 무작위 배정하는 경우와 crossover study와 같이 대상자들이 시간차를 두어 시험군과 대조군에 번갈아 참여하는 경우로 나누어 생각해 볼 수 있다. 많은 경우 시험군과 대조군의 표본크기를 동일하게 산정하게 되는데 이러한 경우 crossover study를 계획하면 요구되는 표본 수를 반으로 줄일 수 있다. 참여하는 대상자 수를 줄인다는 측면에서 crossover study는 매력적으로 보이기도 한다. 하지만 이 경우 대상자의 시험 참여 기간이 너무 길어지게 되므로 중도탈락 가능성이 높아져 시험에 소요되는 시간과 비용이 오히려 더 증가될 수 있을 수 있다. 따라서 시험의 목적과 성격에 적절한 시험의 형태를 정해야 할 것이다.

앞에서 언급한 것처럼 효과평가변수는 연구의 핵심 부분이므로 매우 주의를 기울여 설정해야 한다. 그리고 당연히 표본크기를 산정하기 위해서는 효과평가변수를 기준으로 산정하게 되는데 이를 위해서 효과평가변수를 통해 효과의 척도를 정하는 것이 요구된다. 예를 들어 효과평가변수의 평균, 표준편차, 또는 두 개의 효과평가변수의 상관관계수, 효과평가변수에 대한 관심변수의 회귀계수 또는 오즈비 등이 그 예가 될 수 있다.

효과의 척도가 결정되면 어느 정도의 효과 차이를 의미 있는 차이인지 다르게 말하면 효과 추정량의 최대허용오차를 결정해야 한다. 최대허용오차의 결정은 유사선행연구를 통해 결정하는 것이 일반적이다.

표본크기 산정을 위해 연구자가 준비해야 할 또 다른 사항으로 효과평가변수의 산포도가 있다. 산포도에 대한 척도는 표준편차(분산), 범위(range) 등이 있으며 유사선행연구를 통해 산포도를 결정한다. 최대허용오차와 더불어 산포도가 유사선행연구로부터 확인될 수 있다면 표본크기 산정은 비교적 쉽게 이루어진다. 하지만 유사선행연구를 찾기가 힘든 경우가 대부분이어서 예비조사가 요구된다. 예비조사는 연구 계획과 동일하게 시험을 실시하지만 매우 소규모로 실행이 된다. 예비조사는 본격적인 시험을 실시하기 전에 표본크기 산정에 요구

되는 여러 정보 즉 최대허용오차, 산포도, 중도탈락률을 제공해 줄 수 있다. 뿐만 아니라 연구 계획의 수정 보완 사항도 점검할 수 있는 좋은 기회를 제공해 주기도 한다. 비록 추가적인 비용, 시간, 노력이 요구되기는 하지만 종종 예비조사를 통해 연구에 소요되는 비용, 시간, 노력을 줄여주기도 하므로 가능하다면 예비조사를 실시하여 성공적인 임상시험을 계획하기를 권장한다.

유의수준과 검정력은 연구마다 모두 동일해야 할 이유는 없지만 대부분 유의수준은 0.05로 검정력은 0.7 또는 0.8을 지정해 사용한다. 유의수준은 통계 용어로 표현하면 대립가설에 해당되는 연구자의 주장이 옳지 않은 경우에 적절치 않게 표본이 추출되어 연구자의 주장이 옳다고 판정할 확률에 해당된다. 제1종 오류라고 하는 이러한 오류는 임상시험에서 가급적 발생하지 않기를 기대하므로 그 발생 확률은 매우 작아야 한다. 일반적으로 0.05 정도면 적절한 것으로 받아들여지고 있다. 검정력은 연구자의 주장이 맞은 경우에 임상시험에서 그 주장이 옳음을 지지하는 증거를 확보할 가능성을 의미한다. 유의수준과 달리 검정력은 가급적 높게 책정하는 것이 좋겠지만 이럴 경우 요구되는 표본크기가 많아지는 문제점이 있다.

이상과 같은 정보들이 결정되면 표본크기를 산정하는 공식을 통해 최소표본크기를 계산해 낼 수 있다. 여러 상황에 따라 다양한 산정 공식이 있으므로 적절한 공식을 이용하는 데 주의를 기울여야 하겠다(9). 산정공식을 통해 계산된 표본 수는 임상시험을 실시하여 최대허용 오차만큼의 효과 차이가 연구 가설을 증명하기에 충분한 최소표본 크기를 의미하는 것으로 중도탈락 등의 이유로 자료의 결측이 생기면 추가적으로 자료를 더 수집해야 한다.

4. 통계모형

통계분석은 적절한 통계모형을 미리 설정한 후 이루어진다. 효과평가변수가 연속형인 경우 효과 비교는 평균, 중위수 또는 표준편차를 통해 이루어진다. 효과평가변수에 영향을 줄 관심변수와 교란변수가 없는 경우에는 평균 비교는 모수적 검정법인 t 검정으로, 중위수의 비교는 비모수적 검정법인 wilcoxon 검정, 그리고 표준편차의 비교는 χ^2 검정으로 이루어진다.

연속형 효과평가변수에 영향을 주는 관심변수와 교란변수가 있는 경우에는 회귀모형, ANOVA 모형, ANCOVA 모형을 이용하여 분석이 이루어진다. 회귀모형은 관심변수와 교란변수 모두 연속형 자료인 경우에 사용

하는 모형이다. 관심변수와 효과평가변수 간의 관계를 관심변수에 대응하는 회귀계수의 t 검정을 통해 유의성을 확인하는 것으로 밝힌다. 만일 관심변수와 교란변수들 모두가 범주형인 경우 ANOVA 모형을 이용한 분석을 실시한다. 회귀모형에서와 유사하게 관심변수에 대응되는 F 검정의 결과로 그 유의성을 확인한다. 마지막으로 관심변수와 교란변수들이 연속형과 범주형이 혼합된 경우라며 ANCOVA 모형을 이용한 분석을 실시한다. 이 세 가지 모형들 중에서 ANCOVA 모형은 회귀모형이나 ANOVA 모형을 다 아우르는 모형으로 볼 수 있다. 참고로 이론적인 측면에서 위 세 모형은 별다른 차이는 없다. 다만 그 결과에 대한 해석에 주의를 기울일 필요가 있다. 예를 들어 범주형 관심변수로 인종을 생각해 보자. 만일 연구에 참여한 시험 대상자의 인종은 아시아, 유럽, 아프리카 세 부류로 나눌 수 있다고 하자. 이 경우 인종에 관한 검정 결과는 t 검정과 F 검정 결과를 확인할 수 있다. F 검정은 세 인종에 따른 효과평가변수 값에 유의한 차이가 있는지를 확인한 결과를 제시해 준다. 하지만 t 검정은 2가지의 검정 결과를 제시해 준다. 예를 들면 인종의 reference level이 아시아라면 아시아와 유럽 간의 비교 결과와 아시아와 아프리카 간의 비교 결과이다. 아시아와 유럽 간에 그리고 아시아와 아프리카 간의 유의한 차이가 있으면 세 인종 간에는 유의한 차이가 있다고 할 수 있는가에 대한 질문을 생각해 보자. 얼핏 유의한 차이가 있다라고 단정지을 수 있을 듯 하지만 실제로 그렇지 않을 수도 있다. 그 이유는 유럽과 아프리카 간에는 유의한 차이에 대한 검정 결과가 없기 때문이다. 이러한 한계로 인해 효과평가변수에 대한 범주형 관심변수의 효과는 F 검정으로 확인한다. 만일 F 검정 결과 유의한 효과가 있는 것으로 확인이 되면 비로소 세 인종들 중 두 인종들 간의 차별성을 t 검정을 통해 구체적으로 확인하게 된다. 물론 F 검정의 결과 유의한 효과가 발견되지 않았다면 인종의 효과를 더 이상 논하는 것은 큰 의미가 없게 된다.

효과평가변수가 범주형인 경우는 특정 범주의 비율을 비교의 척도로 사용한다. 예를 들면, 유병률을 생각하는 경우 효과평가변수는 병에 걸린 결과가 되고 범주는 YES와 NO이며 특정 범주는 YES가 된다. 시험군과 대조군 간의 비교는 각 군별 유병률을 z 검정으로 비교한다. 유병률에 영향을 주는 요인 즉 관심변수를 연구하는 경우 유병률 대신 오즈비를 로지스틱 회귀모형을 이용하여 비교한다.

효과평가변수가 생존시간인 경우 연속형이지만 평균

또는 표준편차의 비교보다는 Cox 모형 등을 이용한 생존분석을 통해 효과를 검증한다. 효과에 대한 비교는 hazard ratio를 이용하여 이루어진다. 여기서 hazard는 일정 기간 동안의 사망률이 아닌 순간 사망률 정도로 이해될 수 있다. 생존분석의 특징은 생존시간에서 중도절단자료(censored data)가 많이 발생한다는 점이다. 예를 들어 1월 1일부터 6월 30일까지 실시하는 임상시험에서 시험 대상자의 생존시간을 효과평가변수로 삼는 경우를 생각해 보자. 시험 기간 중간에 사망 이외의 이유로 시험에서 탈락하는 경우 그 시점까지의 생존기간은 중도절단자료로 처리한다. 주의할 사항은 시험 기간 중에 환자가 중간에 퇴원하고 이후에 환자의 사망시기를 나중에 확인한 경우 생존시간은 사망시점까지가 아니고 퇴원시점까지이며 중도절단자료로 처리한다. 그 이유는 퇴원으로 연구환경과 다른 환경인 놓인 대상자의 생존시간은 편이된 결과를 낳을 수 있기 때문이다.

결 론

임상시험은 인간에게 도움이 될 약물 또는 치료 방법의 개발에 필수적인 절차가 되고 있다. 임상시험의 목적은 매우 간단하지만 이를 실시하는 것은 매우 어려운 부분이 있다. 임상시험에서 ‘유익한 결과’를 도출하기 위해서 연구계획을 매우 정교하게 세워야 함은 당연하다. 이를 위해 선행연구에 대한 심층적인 문헌고찰이 이루어져야겠고 연구자 간의 객관적인 시각을 잃지 않도록 노력해야 한다. 이러한 노력의 일환으로 통계를 적극적으로 활용하기를 권한다.

Ederer(3)는 성공적인 임상시험을 위해서 통계학자는 임상시험 전반에 대한 학습을 받을 준비가 되어 있으며 공동연구자에게 통계 관련 사항의 자문을 제공한다고 강조하였다. 임상시험을 계획하는 연구자로서 30년 전의 통계학자의 음미해 볼만한 구절이라 생각한다.

요 약

임상시험은 인간을 대상으로 약물 또는 치료법의 효과를 검증하는 것을 목적으로 하고 있다. 성공적인 임상시험을 위해서는 단순한 자료분석에만 통계의 이용을 제한하지 않고 다양한 영역으로 활용의 폭을 넓히는 것이 필요하다. 연구계획단계에서부터 구체적이고 체계적으로 통계의 활용을 고려하기 위해 효과에 대한 정의, 적절한 표본크기 산정, 통계분석 방법 등 전반적인 통계의 응용을 고찰한다.

중심단어: 임상시험, 표본크기, 통계학

참 고 문 헌

1. Willcox WF. The founder of statistics. Rev Int Stat Ins 1938; 5(4):321-8.
2. Wikipedia.org [Internet]. Statistics; c2008 [cited 2010 Jan 13]. Available from: http://en.wikipedia.org/wiki/Statistics#cite_note-9.
3. Ederer F. The statistician's role in developing a protocol for a clinical trial. Am Stat 1979;33(3):116-9.
4. Bearman JE. Randomized controlled clinical trial. National Eye Institute workshop for ophthalmologists. Writing the protocol for a clinical trial. Am J Ophthalmol 1975;79(5):775-8.
5. Kaku DA, Lowenstein DH. Emergence of recreational drug abuse as a major risk factor for stroke in young adults. Ann Intern Med 1990;113(11):821-7.
6. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. N Engl J Med 1989;321(3):129-35.
7. Dawson B, Trapp RG. Basic & clinical biostatistics. 3rd ed. New York:McGraw-Hill;2001.
8. Parker RA, Berman NG. Sample size: more than calculations. Am Stat 2003;57(3):166-70.
9. Oh HB, Ahn HY, Lee KJ, Hwang SH, Choi SE, Lee HJ. Clinical evaluation of diagnostic medicines. Seoul:Korea Medical Book; 2007.