

한국심리학회지 : 임상
The Korean Journal of Clinical Psychology
2015, Vol. 34, No. 4 1101-1112

〈특별호〉

치료효과 연구설계의 통계적 측면*

박 미 라

을지대학교 의과대학
예방의학교실

김 선 우[†]

삼성생명과학연구소
통계지원팀

연구문제에 대한 해답은 신뢰할 수 있고 타당성 있는 결과이어야 하며, 이러한 신뢰성과 타당성은 과학적인 연구설계에서 비롯된다. 정교하고 편의 없는 연구결과를 도출하기 위해서는 연구의 계획에서부터 실행, 분석과 해석에 이르기까지 전 과정에서 구체적인 프로토콜을 가지고 연구를 진행하여야 한다. 특히 연구설계 단계에서의 통계적 고려는 과학적인 연구를 이끌어 나가기 위한 중요한 요소 중 하나이다. 연구설계 단계에서 근본적 결함이 발생하면 연구 진행 과정이나 종료 후에 완전한 보완이 불가능하므로, 사전에 충분한 준비과정이 필요하다. 이에 본 논문에서는 연구의 설계 단계에서 고려해야 할 주요 항목들과 통계적 문제를 예시와 함께 소개하고자 한다.

주요어 : 연구설계, 확률화, 표본크기, 통계분석

* 이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2013R1A1A2062848).

† 교신저자(Corresponding Author) : 김선우 / 삼성생명과학연구소 통계지원팀 / 서울특별시 강남구 일원로 81 / E-mail : seonwoo.kim@samsung.com

서 론

일반적으로 연구의 진행은 연구설계, 실험이나 조사를 통한 데이터의 수집, 데이터의 분석과 결과의 해석, 그리고 논문 작성 등의 순서로 진행된다. 연구를 성공적으로 마치기 위해서는 이러한 연구의 모든 단계에서 통계적 사고가 필요하다. 특히, 각 단계의 품질은 이전 단계가 얼마나 과학적으로, 그리고 효율적으로 수행되었느냐에 큰 영향을 받기 때문에, 연구의 첫 단계인 연구설계를 과학적이고 효과적으로 수립하는 것은 매우 중요하다.

그럼에도 불구하고, 아직도 많은 연구에서 통계적 뒷받침이 없이 잘못된 연구설계를 바탕으로 연구를 진행하여, 연구의 질 저하와 비용 낭비를 초래하고 있다. 세계적으로 권위 있는 30개 의학전문학술지에 발표된 4,235편의 논문을 대상으로 연구의 과학성, 연구설계, 통계적 방법, 그리고 논문에서의 연구 결과 제시 등에 관한 검토를 했을 때 적절하다고 평가된 논문은 20% 정도에 불과하다는 연구결과가 있다(Williamson, Goldschmidt & Colton, 1986). 또한 이들 중 일부를 선택하여 조사한 결과, 확률화, 맹검 및 통계분석방법 등 불충분하고 계획이 잘못된 연구의 80%가 통계적으로 유의한 결과(이른바 positive findings)를 얻었으나, 제대로 계획된 연구의 경우에는 단지 25%만이 유의한 결과를 얻을 수 있었다고 한다. 즉 원하는 결과는 오히려 잘못 계획된 연구에서 더 많이 얻어지고 있다는 결론이다. 제대로 계획되지 않은 연구의 결과의 발표는 과학성과 타당성의 결여로 그 악영향은 심각하다. 특히 연구가 사람을 대상으로 하는 치료효과성 연구일 때 그 심각성은 더 할 것이다.

연구자가 알고 싶은 연구문제는 단순하게 기술할 수 있지만, 그에 대한 해답을 위한 설계는 단순하지 않을 수도 있다. 그러나, 몇 가지 중요한 사항들을 점검함으로써 과학적인 연구에 가까이 갈 수 있을 것이다. 본 연구에서는 임상심리연구에서 의미 있는 결과를 도출하기 위해 연구자들이 연구설계 단계에서 기본적으로 고려해야 할 항목과 통계적 문제들을 간단한 예와 함께 제시하고자 한다.

연구설계에서의 통계적 고려

연구개요와 연구계획서

연구계획을 잘 수립하기 위한 좋은 방법 중 하나는 우선 하나의 문장으로 연구문제를 써보는 것이다. 이를 중심으로 하여 개요(outline), 연구계획서(study protocol), 실행지침서(operations manual)로 점차 더 자세하게 작성해나가면 된다(Huley, Cumming, Browner, Grady & Newand, 2001). 개요에서는 연구에 필요한 요소를 연구의 흐름에 따라 순차적으로 채우면 되는데, 1쪽 정도로 작성한다. 여기에는 연구에 필요한 항목이 모두 포함되어 있으므로 체크리스트 역할을 할 수 있으며, 주제에 대한 연구자의 구상을 명확히 하는 데 도움이 된다. 또한 같은 분야 내 또는 다른 분야간 공동연구자들간의 소통에도 큰 도움이 될 수 있다. 다음의 표 1은 Hulley et al.(2001)이 추천하는 개요양식을 토대로 재작성한 것이다.

다음 단계로 이를 연구계획서나 실행지침서로 확장한다. 연구계획서는 연구개요의 내용을 확장하여 보다 구체적으로 설명한 것으로 연구를 계획하고 연구비를 신청하는데 사용될

표 1. 연구개요의 양식

요소	세부요소
1. 제목	
2. 연구문제	
3. 연구배경과 의의	
4. 연구설계	<ul style="list-style-type: none"> - 관측연구/실험연구 - 전향(prospective)/후향(retrospective) - 종단(longitudinal)/횡단(cross-sectional)
5. 연구대상	<ul style="list-style-type: none"> - 포함기준 - 표본추출방법
6. 변수	<ul style="list-style-type: none"> - 설명변수 (predictor) - 결과변수 (outcome) - 교란변수 (confounder)
7. 통계적 문제	<ul style="list-style-type: none"> - 가설 - 표본크기 - 분석방법

수 있다. 실행지침서는 연구의 품질관리를 위해 연구실행을 통일된 방식으로 진행 할 수 있도록 특정 절차에 대해 지시나 질의사항 등을 적어 놓은 것을 말한다.

연구문제

연구문제(research question)는 연구를 진행하는 이유이자 동기로, 연구의 안내자이자 중심이다. 따라서 어떤 연구문제를 해결하고자 한다면 그 문제가 무엇인지를 구체적이고 전반적으로 알아야 하고, 연구를 통하여 얻고자 하는 목적을 분명히 해야 한다는 말이다. 이는 자명하고 간단한 일 같지만 실제로 상당한 고민과 시간이 필요하다. 연구문제를 지나치게 총괄적으로 서술하거나 지나치게 간략하게 서술하면 연구를 진행하는데 많은 어려움에 봉착하게 된다. 따라서 연구문제를 면밀하게

서술하는 것은 연구의 계획과 진행에 큰 도움이 된다. 그렇다면 연구문제를 어떻게 서술하는 것이 좋을까? 첫째, 가능하다면 둘 혹은 그 이상의 변수간의 관계로 표현함이 좋다. 둘째, 연구문제는 질문형식으로 분명하고 명확하게 서술되는 편이 편리하다.

예를 들어 ‘우울 증상 치료에 대한 심리치료 효과’라고 하는 것보다는, ‘우울 증상이 있는 청소년들에 대해 심리치료가 기존 약물투여에 비해 더 효과적인가?’ 또는 ‘청소년 우울증상 치료에 심리치료가 기존 약물투여만큼 효과적인가?’와 같이 서술하는 것이 좋다. 이를 통해, 우선 연구대상의 범위가 우울 증상이 있는 청소년임을 알 수 있다. 그리고 새로운 치료법인 심리치료방법에서 기대하는 효과는 전자의 경우는 더 우월한 치료 효과를, 후자의 경우는 동등한 치료효과를 기대하는 것임을 보다 명확히 할 수 있게 한다.

연구대상

모집단에는 목표모집단과 표본모집단이 있다. 목표모집단이란 연구의 대상이 되는 모집단이며, 표본모집단이란 연구의 여건 등을 이유로 연구자가 취급한 모집단이다. 이 두 모집단의 속성의 차이는 클 수도 작을 수도 있다. 그러나 이 두 모집단을 구분하여 언급해 놓는 것이 필요하다. 왜냐하면 연구에 사용되는 표본은 표본모집단으로부터 추출하기 때문에, 연구결과의 적용 대상은 목표모집단이 아니라 표본모집단이 되기 때문이다. 또한 이 두 모집단을 구분하여 언급함으로써 연구의 범위가 구체화 된다. 이를 위해서는 표본을 얻기 위한 모집단의 규모, 구조, 특징을 살펴 보아야 할 것이다. 예를 들어, 청소년 우울 증상에 대한 심리치료의 효과 연구의 대상으로 목표모집단은 대한민국 13세 이상 18세 이하인 청소년이나, 연구 편의상 특별시와 광역시에 거주하는 13세 이상 18세 이하인 청소년을 대상으로 했다면 이들이 표본모집단이 된다. 따라서 이 연구의 결과는 표본모집단에 대해서만 적용가능하다.

연구문제에 적합한 대상자의 인구통계학적, 임상적 특성을 수립하여 구체적인 포함기준을 정할 수 있다. 또한 부적절한 대상에 대한 몇 가지 배제기준을 정할 수 있다. 예를 들어 청소년 우울 증상 연구에서 2014년도에 특별시와 광역시에 거주하는 13세 이상 18세 이하인 건강한 남녀를 대상으로 하며, 언어장애가 있는 자는 제외한다고 정할 수 있다.

표본모집단에서 표본을 뽑는 방법으로는 편의추출(convenience sampling)이나 단순임의 추출(simple random sampling), 집락추출(cluster sampling)이나 층화추출(stratified sampling), 체계

적 추출(systematic sampling) 등의 표본추출법이 있다. 이들 표본추출방법에 대해서는 김호일(2014)이나 성내경(2012b)과 같은 관련서적을 참고하기 바란다.

변수의 선택

연구설계시 결정해야 할 중요한 문제 중 하나는 측정할 결과변수를 선택하는 것이다. 연구자는 보통 연구에서 답을 구하고자 하는 문제들이 아주 많다. 하지만 연구에서 가장 알고 싶은 목적(연구의 1차 목적이라 함)을 연구 설계 단계에서 분명히 설정하고 있어야 한다. 이로부터 연구의 1차가설과 1차 결과변수가 결정되며, 이에 맞게 필요한 표본의 크기도 계산하게 된다.

앞의 우울 증상 치료법을 예로 들어보자. 우울 증상 감소와 자살 충동 감소를 치료 효과로 간주한다면, 이 두 가지를 모두 포함하는 것을 1차 목적으로 할 것인지, 아니면 이 중 하나만을 1차 목적으로 하고 다른 하나는 2차 목적으로 할 것인지 결정한다. 만약 우울 증상 감소를 1차 목적으로 한다면, 증상 감소를 반영하는 변수를 미리 계획해야 하는데, 추상적 개념의 변수가 아니라 측정 가능한 변수로 정의해야 한다. 측정 가능한 변수로 증상 재발 여부, HAM_D(Hamilton rating scale for depression) 점수 변화량 등이 있을 수 있다. 연구자가 치료 효과를 가장 잘 반영하는 변수로서 재발 여부를 선택했다면, 재발 여부가 이 연구의 1차 결과변수가 되고, 치료 효과를 반영하는 다른 측정 변수는 2차 결과변수가 된다.

이 때 한 가지 더 깊게 들어가야 하는 포인트가 있는데, 그것은 재발 여부 측정 시점이

다. 이 시점은 치료 후 1년 또는 마지막 추적 관찰 시점 등이 될 수 있다. 치료 후 1년 시점에서의 재발 여부를 1차 결과변수로 정의하느냐, 마지막 관찰 시점에서의 재발 여부를 1차 결과변수로 정의하느냐에 따라 표본 크기 산출 방법과 적용할 통계분석 방법도 달라진다. 만약 전자를 1차 결과변수로 정의했다면, 연구자는 모든 환자들을 치료 후 1년 시점까지 추적 관찰해야한다. 만약 이 1차 결과변수가 사전에 계획되지 않았다면, 1차 결과변수를 측정하지 못한 연구 대상자가 있을 수 있으며, 이는 표본 크기의 감소 및 편향된 연구 결과를 초래할 수 있게 된다.

분석적 연구에서 두 변수간의 관계를 알고자 한다면 원인으로 생각되는 설명변수(predictor variable)와 결과로 생각되는 결과변수(outcome variable)를 설정한다. 임상연구에서는 보통 중재(intervention)에 따른 결과를 비교하게 되는데, 이 경우 중재여부가 설명변수가 될 수 있다. 이밖에 결과에 영향을 미치면서 이를 변수간의 인과관계를 유추하는데 혼란을 부를 수 있는 교란변수(confounding variable)를 고려한다.

연구가설과 통계적 형식화

가설은 과학적인 연구에서 중요하고도 필수적인 중간 디딤돌이라는 것에 대해서는 의심 할 여지가 없다. 연구문제는 보통 추상적인 속성을 지니고 있으며 직접적으로 검증되지 않기 때문에, 가설의 형태로 축소되지 않으면 과학적으로 해결할 수 없다. 가설이란 연구자가 연구에서 입증하고자 하는 것에 대한 서술이다. 가설은 연구자들로 하여금 자기 자신을 객관적으로 바라볼 수 있게 하기 때문에 연구

를 진행하는 효과적인 수단이 된다. 가설은 연구문제 또는 연구목적과 부합되어야 하고 측정 변수로 기술되어야 하며, 비교 집단이 포함되어야 하고, 연구에서 기대하는 바가 포함되도록 구체적으로 만들어져야 한다. 비교 하려는 변수에 대해 집단간 차이가 있는지를 주장하기 위한 연구(비교연구, comparative study)인지, 또는 비교하려는 변수가 집단간 차이가 없음을 주장하기 위한 연구(동등성 연구, equivalence study)인지 명확히 알 수 있어야 한다. 또한 양측검정(two-sided test)을 해야 하는지 또는 단측검정(one-sided test)을 해야 하는지, 가설검정에 필요한 변수들이 무엇인지 명확해야 한다.

예를 들어 연구목적이 ‘우울 증상이 있는 청소년들에 대해 심리치료가 기존 약물투여방법에 비해 더 효과적인가?’와 같이 서술되었고, 치료 효과를 반영하는 변수로 치료 후 1년 시점에서의 재발 여부가 선택되었다고 하자. 이 때, 기존의 약물투여방법에 비해 심리치료가 더 효과적임을 입증하고자 한다면, ‘심리치료 후 1년 재발율이 약물투여방법에서의 1년 재발율보다 더 낮다.’로 연구가설을 세울 수가 있다. 그러면 여기서 우리는 이 연구는 비교연구이며, 단측검정을 해야 하고, 가설검정에 필요한 변수는 치료 후 1년 재발율임을 분명히 알 수가 있다. 그러나 연구가설이 “우울 증상의 재발방지에 심리치료방법이 기존의 약물투여방법만큼 효과적인가?”라 하면, 이 연구는 동등성 연구가 되므로, “치료 후 1년 재발율이 약물투여방법과 심리치료방법간 차이가 없다.”로 연구가설을 세워야 한다.

가설은 통계적 표현을 이용하여 형식화함으로써 보다 분명하게 표현할 수 있으며 논리전개의 비약을 피할 수 있다. 예를 들면 ‘흡연은

인지능력감소와 연관이 있는가?’라는 연구가설은 ‘흡연군과 비흡연군간의 인지능력감소비율이 차이가 있는가?’라는 통계적 가설로 바꿀 수 있고, ‘두 치료방법의 1년 재발율의 차이가 남녀에 따라 다른가?’라는 연구가설은 ‘두 치료방법과 성별간의 1년 재발 여부에 대한 교호작용효과(interaction effect)가 있는가?’라는 통계적 가설로 바꾸면 연구가설을 검정할 통계적 기법이 보다 명확하게 선택될 수 있다.

확률화

사람을 대상으로 진행되는 임상연구에서 얻어지는 자료는 우리가 관리 또는 조정할 수 없는 여러 가지 요인 때문에 편의(bias)가 발생할 여지가 크다. 편의를 줄이는 중요한 방법으로 확률화(randomization)를 통한 연구대상자의 배정이 있다. 예컨대, 약물치료방법과 심리치료방법을 비교하는 연구에서, 연구 대상자들을 두 군(약물치료군, 심리치료군)으로 나누어 배정하는 방법을 생각해보자. 이때 연구대상자들은 치료방법 외에 다른 요인들에는 차이가 없어야 좋을 것이다. 확률화를 통해 대상자를 랜덤하게 두 군 중 하나에 할당하게 되면, 치료방법 외에 결과변수에 영향을 미칠 수 있는 나이나 성별, 기타 여러 가지 예후인자들이 두 군 간에 동일하게 분포되어 있을 것이라고 가정할 수 있게 된다. 따라서 치료효과의 차이가 치료 방법에만 기인하는 것으로 해석이 가능하다는 큰 장점을 가질 수가 있다.

그러나 항상 확률화가 가능한 것은 아니며, 연구 진행 방법이나 윤리적인 문제로 인해 불가능한 경우가 있다. 예를 들어 많은 연구 대상자들이 심리치료를 원하게 되면, 약물치료

군에 배정된 증상이 경미한 연구 대상자는 연구 참여를 철회할 가능성이 높아서, 약물치료군에는 증상이 심한 대상자만 남게 되어 연구 결과에 편의를 불러일으킬 수가 있고, 연구 대상자 모집에 어려움을 겪게 된다. 또한, 증상이 심한 연구 대상자들이 많은 경우 심리치료군에 임의로 배정하는 것은 윤리적 문제의 소지가 있게 된다.

또한 확률화에 의해 연구 대상자들을 두 군으로 나누었다고 해서, 치료법 이외의 모든 인자들이 반드시 동일하게 되는 것은 아니다. 수십명의 표본수로 진행되는 연구가 많은데, 이렇게 유한한 표본에서는 두 군간 차이가 나타나는 인자가 나올 수가 있다. 따라서 통계분석 단계에서 연구 결과에 영향을 미칠 수 있는 주요 인자들이 두 군간 차이가 있는지 반드시 확인하고, 유의한 차이를 보이는 인자에 대해서는 통계분석 과정에서 사후 보정하는 작업을 거쳐야한다.

확률화 방법으로 단순확률화(simple randomization) 외에도 블럭확률화(block randomization), 충화확률화(stratified randomization) 등이 있다. 자세한 내용은 관련문헌을 참고하기 바란다(Altman, 1991; 이재원, 박미라, 유한나 2011). 인터넷상에서 무료로 제공되는 확률화할당 프로그램을 이용하면 확률화 작업을 수행하기가 수월하다 (<http://www.randomization.com/>).

짝짓기

앞서 언급했듯이 확률화가 모든 인자에 대해 항상 균형(balance)을 보장하는 것은 아니다. 만약 연구 결과에 아주 큰 영향을 미치는 인자가 있고 이 인자에 대해서 반드시 비교 그룹간 균형을 이루어야한다면, 이 인자에 의해

짝짓기(matching)를 수행하는 것이 좋다. 약물치료방법과 심리치료방법을 비교하는 연구에서, 치료의 효과가 연령에 큰 영향을 받는다면, 연령으로 짝짓기를 하는 것이 좋다. 동일 연령의 연구 대상자 두 명 중 랜덤하게 한 명은 약물치료방법, 다른 한 명은 심리치료방법으로 배정하는 것이다. 동일 연령에 대해 짝짓기를 진행하는 경우, 연구대상자 모집에 어려움이 있다면, 연령을 범주화하여 연령군내에서 짝짓기를 진행할 수도 있다. 예컨대 14 세인 사람의 짝을 11-15세 사이에서 찾는 방법이다. 1:1 짝짓기가 어려울 때에는 그룹짝짓기(group matching)를 고려할 수도 있다. 이 방법은 예컨대 한 그룹의 여성비율이 30%라면 다른 집단도 30%가 되도록 맞추는 것이다.

짝짓기 방법은 확률화 방법에 비해 해당 요인에 대해 훨씬 균형이 잘 이루어지며, 대부분 적은 연구 대상자 수로 진행할 수 있다는 장점이 있으나, 짝짓기 대상을 모집하기가 어려운 경우 시간이 많이 소요될 수가 있다.

표본 크기

표본 크기를 산출한 근거를 제시하는 것은 이제 모든 논문에서 필수적이 되고 있다. 표본 크기를 결정하는 문제는 연구의 검정력(power)을 결정하는 문제와 관련되며, 특히 기대한 정도의 차이가 발견되지 않은 연구에서 검정력의 언급은 필수적이다. 검정력이란 연구자가 제기한 가설이 옳을 때, 기대한 결과를 얻을 수 있는 가능성을 말한다.

표본 크기가 너무 크면, 검정력이 너무 높아져서 임상적 또는 실제적으로 의미가 없는 작은 차이도 통계적으로는 유의하게 되어 통계적 해석과 임상적 해석이 서로 부합되지 않

을 수 있다. 또한 불필요하게 연구비를 낭비하게 되며, 확률화 임상시험의 경우 불필요한 수의 환자들이 효과가 낮은 치료 또는 아직 검증되지 않은 치료를 받아야하는 윤리적인 문제도 발생시킨다. 반대로 표본의 크기가 작으면, 검정력이 낮아져서 실제로는 차이가 있는데도 불구하고 통계적 유의성을 검출하지 못하는 경우가 많다. 또한 차이가 검출된 연구결과라 할지라도 그 신뢰성은 낮게 된다.

적절한 표본 크기는 1차 결과변수와 연구의 유형에 따라 달라지며, 통계분석방법, 유효크기(effect size), 유의수준, 검정력 등에 따라서도 달라지게 된다. 1차 결과변수가 동일한 연구라도 탐색연구인지 확증연구인지, 확증연구 중에서도 비교연구인지 동등성연구인지에 따라 필요로 하는 표본의 크기는 매우 달라질 수 있으므로, 연구에 적절한 표본 크기 산출방법을 찾아야 한다. 여기서 탐색연구(exploratory study)란 연구가설이 확립되지 않은 상태에서 기술통계량으로 평가변수를 탐색하는 것이 목적인 연구로서 주로 치료 효과의 추정치와 그의 신뢰구간을 구하게 된다. 확증 연구(confirmatory study)는 연구가설이 확립되어 통계적 가설검정절차를 통한 연구가설의 각각 여부를 확인하여 치료 효과 여부에 대한 결론을 내리는 연구를 말한다. 표 2는 몇 가지의 경우에 대한 표본 크기 산출 계산식의 예를 보여준다.

예를 들어 심리치료가 어느 정도 효과가 있는지를 알아보는 연구에서, 일차결과변수를 치료 후 1년 시점에서의 재발 여부라 할 때, 이 연구는 재발율과 재발율의 95% 신뢰구간을 산출하는 것을 일차목적으로 하는 탐색연구가 될 것이며, 이 연구 결과 어느 정도의 효과가 있다고 탐색이 되면, 기존 치료방법과

표 2. 표본 크기 산출 계산식의 예

연구유형	결과변수	표본 크기 산출 계산식	목적
탐색 연구	연속형	$N = 4Z_{\alpha/2}^2 S^2 / W^2$	평균과 신뢰구간 추정
	이분형	$N = 4Z_{\alpha/2}^2 P(1-P) / W^2$	비율과 신뢰구간 추정
확증 연구	연속형	$N = [2(Z_{\alpha/2} + Z_{\beta})^2 S^2] / d^2$ 각 군당 필요한 표본 크기	두 군의 평균 비교, 1:1 할당, 양측검정
	이분형	$N = \frac{(Z_{\alpha/2} \sqrt{2P(1-P)} + Z_{\beta} \sqrt{P_1(1-P_1) + P_2(1-P_2)})^2}{(P_1 - P_2)^2}$ 각 군당 필요한 표본 크기	두 군의 비율 비교, 1:1 할당, 양측검정

여기서 S : 표준편차, W : 구간의 원하는 너비, P : 기대되는 비율, P_1 : 대조 집단에서의 비율, P_2 : 실험집단에서의 비율, $\bar{P} = (P_1 + P_2)/2$, d : 유효크기, $Z_{\alpha/2}$ 와 Z_{β} : 각각 $\alpha/2$ 와 β 에 해당하는 표준정규분포의 임계치. 단측검정의 경우 $\alpha/2$ 대신 α 를 사용함

비교하는 확증연구가 다음 연구로 진행될 것이다. 이 탐색연구에서, 표본모집단에서 심리치료 후 1년째 재발율이 20%로 기대되고, 95% 신뢰구간의 폭을 ±5%로 하고자 한다면, 246명의 우울증 청소년이 필요하다.

$$N = 4Z_{\alpha/2}^2 P(1-P) / W^2 \\ = 4 * (1.96)^2 * 0.2 * (1-0.2) / (0.1)^2 \\ = 246$$

그러나 기존 약물치료방법과 심리치료방법이 효과의 차이를 입증하는 확증연구에서는, 기존 약물치료방법에서 1년 재발율이 30%이고, 심리치료방법을 사용하면 최소한 20%이하로 감소할 것을 기대한다면, 단측검정으로 진행하고자 할 때, 유의수준 5%, 검정력 80%에서, 각 군당 231명이 필요하다.

$$N = \frac{(Z_{\alpha} \sqrt{2P(1-P)} + Z_{\beta} \sqrt{P_1(1-P_1) + P_2(1-P_2)})^2}{(P_1 - P_2)^2} \\ = \frac{(1.645 \sqrt{2*0.25*0.75} + 0.84 \sqrt{0.3*0.7 + 0.2*0.8})^2}{(0.3 - 0.2)^2} \\ = 231$$

그러나 약물치료의 부작용 때문에 심리치료를 대안으로 생각해내고, 심리치료의 효과가 약물치료의 효과와 동등함을 보이는 것이 연구의 1차목적인 확증연구라고 하자. 약물치료 후 1년째 재발율이 30%이고, 심리치료의 경우 이 재발율이 최대 5%이상 차이가 나지 않으면 효과가 동등하다고 할 때, 유의수준 5%, 검정력 80%하에서 각 군당 1361명이 필요하다.

이 밖에 다양한 경우의 표본 크기 산출방법에 대해서는 이재원 등(2011)을 참고하기 바란다. PASS나 G*POWER를 비롯하여 연구

대상자수를 계산해주는 다양한 유상/무상의 소프트웨어들이 많이 있다(<http://www.ncss.com/software/pass/>, <http://www.gpower.hhu.de/> 등).

통계분석방법의 선택

일반적으로 통계분석방법은 데이터 수집 이후 결정한다고 생각되고 있으나, 이는 잘못된 인식이다. 연구에서 사용할 통계적 기법은 연구설계 단계에서 설정된 연구가설 그리고 평가 변수를 통계적으로 검정하고 분석하는 기법이며, 이 기법들은 연구설계 단계에서 미리 계획되고 실제 분석 단계에서는 계획된 방법에 따라 진행되어야 한다. 그렇지 않으면, 연구 자료를 보고 유리한 통계적 기법을 선택하거나, 원하는 결과가 나올 때까지 분석을 계속하는 등의 소지가 많게 되는데, 이렇게 해서 도출된 유의한 결과는 재현성 가능성이 매우 낮기 때문에 치료 현장에 적용하는 것은 매우 위험하다.

특히, 결측치 처리 방법이나 부분군 분석 방법에서 이러한 경우가 종종 발생한다. 자료를 보고 원하는 결과가 나올 수 있도록 여러 가지 결측치 처리 방법 중 유리한 방법을 선택한다거나, 원하는 결과가 나올 때까지 계획되지 않은 부분군 분석을 수행하는 경우이다. 약물치료효과와 심리치료효과를 비교하는 연구에서 정해진 시점에서 HAM_D 점수가 측정이 되지 않은 환자에 대해서, 해당 결측값을 어떻게 처리할 것인지를 미리 계획하여 연구계획서에 기술하여야 한다. 또한, 우울 증상이 상대적으로 낮은 그룹과 높은 그룹 각각에서도 심리치료가 약물치료보다 효과가 있는지를 알아보는 것이 임상적으로 관심이 있다면 이를 연구계획서에 반영하고 적합한 통계분석

방법을 기술해놓아야 한다.

적절한 통계분석방법을 계획하기 위해서는 분석할 변수의 성질, 예를 들어 정규분포를 따르는가, 그룹간 분산이 동일한가 등에 대한 사전 정보가 필요하다. 이러한 정보는 관련 문헌이나 이전 단계의 연구, 즉 탐색연구의 경우 pilot study, 확증연구의 경우는 탐색연구에서 얻을 수 있으므로, pilot study, 탐색연구, 확증연구의 순서로 연구를 진행하는 것이 적절하다. 그렇지 않으면, 잘못된 정보로 인하여 부적절한 통계분석방법이 계획될 수 있다.

부분군 분석

부분군 분석은 전체 자료에서 분석한 동일한 내용을 부분군으로 나누어 분석하는 것으로, 일부 특정군에서도 치료효과가 있는지를 입증하기 위해 사용된다. 부분군 분석은 부분군간 표본 크기가 다르거나, 비교할 변수의 변이가 다르거나, 또는 비교 변수에 영향을 줄 수 있는 다른 변수들의 불균형으로 인해, 특정 부분군에서 치료효과가 높거나 낮게 나타나는 왜곡 현상이 발생할 수 있다. 따라서, 부분군 분석이 필요할 경우, 연구설계 단계에서 부분군을 층으로 하는 층화학률화를 적용하는 것이 가장 좋다. 표본 크기는 통계적 방법으로 사후 보정이 불가능하고, 다른 변수들의 불균형도 통계적 방법으로 완전 보정 또는 제어가 어렵기 때문이다. 예를 들어, 평가 변수가 치료 후 6주째 HAM_D 점수의 변화량이고, 이 변화량이 평균이 0인지를 남녀별로 검정한다고 하자(남자 10명, 여자 20명). 평균 변화량이 남녀 모두에서 10점이고, 표준편차도 남녀 모두에서 15점임에도 불구하고, 남자와 여자에서 p-value는 각각 0.06과 0.01로 서로 다

른 결과를 준다. 여기서, 남자군이 여자군보다 고연령층이 많다면, 동일한 표본 크기와 변이를 갖더라도, 남자군에서 HAM_D 개선이 덜 할 수 있다. 총화학률화방법 적용이 현실적으로 어려울 경우, 여자군과 남자군을 연령에 대해 매칭하여 두 군간 연령의 균형을 맞춘 후 분석을 수행할 수 있다.

논 의

지금까지 치료연구의 설계시에 고려해야 할 항목들을 살펴보았다. 연구설계에 있어서 통계적 측면은 연구의 구조를 비롯하여 데이터를 수집하고, 측정변수를 선택하고, 필요한 표본 크기를 산출하는 것까지 모든 과정에 관련되어 있음을 알 수 있다. 이 밖에 연구설계시 고려해야 할 세부항목으로 변수의 측정(measurement), 표본의 대표성, 그룹선택의 기준 등이 있다. 변수의 유형(연속형, 범주형 등)을 선택하는 것 외에도, 정도(precision)를 높이고 편의(bias)를 줄이는 방법을 고려하는 것이 좋다. 예컨대, 정확한 측정을 위해 관찰자를 훈련시키고 둘 이상의 관찰자를 두거나, 측정도구를 정비하여 오차를 줄이고, 측정을 반복하여 정도를 높일 수 있다. 대부분 연구로부터 얻은 결과를 모집단으로 확장하고 싶어 한다. 이러한 측면에서 특히 주의해야 할 점으로 첫째, 표본이 관심 있는 모집단을 대표할 수 있어야 하고, 둘째, 비교할 그룹은 직접 비교할 요인을 제외한 나머지 부분에서는 최대한 비슷해야 한다는 것이다(Altman, 1991). 또한 비교할 그룹을 어떻게 정할 것인지도 중요한 문제이다(Cutis, 1986).

연구설계를 선택하는데 있어서 만약 윤리적

으로나 논리적으로 모두 가능하다면 관측연구보다는 실험연구를 하는 것이 더 선호된다. 실험연구가 가능하지 않고 관측연구를 해야 한다면 코호트연구, 사례-대조연구, 횡단연구의 장단점을 살펴서 적절한 방법을 선택해야 한다(Indrayan, 2013). 또한 독립표본이나 짹지은 표본, 반복측정된 표본 중 어느 것을 사용할 지도 결정해야 한다(Coolican, 2013). 실험연구를 진행하는 경우에도 일원배치, 요인배치, 교차계획 등의 다양한 실험방법 중 적절한 것을 선택해야 한다(성내경, 2012a; 이우선, 정형철, 박만식, 2014). 그러나 여러 가지 고려해야 할 사항이 많으므로 적절한 설계방법을 선택하는 것은 쉽지 않다. 좋은 방법 중 하나는 연구의 초기부터 통계전문가의 도움을 받는 것이다. 최근 국내외적으로 권위 있는 학술지에는 대부분 통계학자들이 편집진으로 참여하여 투고되는 논문들을 심사하고 있다. 실제로 투고된 논문들의 일부 심사결과가 연구설계 단계에서의 문제점들이 지적되어 대응이 곤란하거나 아예 불가능한 경우도 많이 발생한다. 연구설계에서부터 통계전문가와의 협동 연구를 진행하면, 향후 분석이나 논문수정에서의 대응은 물론이고 연구시작시부터 표본추출방법이나 대조군의 설정, 실험계획법의 선택시에 도움을 받을 수 있고, 적절한 표본 크기 산출을 요청할 수 있다.

잘 계획된 연구의 결과는 보다 과학적이고 타당하며 신뢰할만한 것이 될 수 있다. 본고에서 언급한 사항을 검토하는 절차를 거침으로써 연구가 갖고 있는 제약점을 사전에 파악할 수 있을 것이다. 그리고 이러한 항목들은 대부분의 저널에 명확히 언급되어야하는 항목이기도 하다. 본고에서 다룬 것 외에 보다 다양한 상황에 대한 주의사항들은 Bailar &

Hoaglin(2012), Coolican(2013)이나 Hulley et al.,(2001), Indrayan(2013) 등을 참고하기 바란다.

참고문헌

- 김호일 (2014). 표본추출방법론. 경문사.
- 성내경 (2012a). 실험설계와 분석 제2판. 자유아카데미.
- 성내경 (2012b). 표본조사방법론. 자유아카데미.
- 이우선, 정형철, 박만식. (2014). 최신 실험설계 와 분석. 민영사.
- 이재원, 박미라, 유한나. (2011). 생명과학연구를 위한 통계적 방법론(개정판). 자유아카데미.
- Altman, D. (1991). Practical statistics for medical research. Chapman & Hall.
- Bailar, J. C and Hoaglin, D. C. (2012). Medical uses of statistics, 3rd Edition, Wiley.
- Coolican, H. (2013). Research Methods and Statistics in Psychology, 5th Edition. Routledge.
- Cutis, L. M. (1986). Clinical trials. New York: Oxford University Press.
- Hulley, S., Cumming, S., Browner, W., Grady, D., and Newand, T. (2001). Designing clinical research, 3rd Edition, Wolters Kluwer.
- Indrayan, A. (2013). Medical biostatistics. Chapman & Hall.
- Robson. C. (1999). Experiment, Design and statistics in psychology, 3rd Edition, Penguin UK.
- Williamson, J. W., Goldschmidt, P. G., and Colton, T. (1986). The quality of medical literature: An analysis of validation assessments. Medical uses of statistics. In: Bailar JC, Mosteller F, ed. Massachusetts Medical Society. Wiley.

원고접수일 : 2015. 10. 07.

수정원고접수일 : 2015. 10. 24.

제재결정일 : 2015. 10. 29.

〈Special Section〉

Statistical aspects for the design of the study for treatment effect

Mira Park¹⁾

Seonwoo Kim²⁾

¹⁾Department of Preventive Medicine, Eulji University

²⁾Biostatistics Team, Samsung Biomedical Research Institute

Results from the research question should be reliable and valid, and depend on the scientific design of the study. For results with high precision and no bias, the study should proceed carefully according to a detailed protocol at each stage, including study conducting, data analysis, interpretation, etc., throughout the study period. Statistical design is required to lead the scientific design. Since fallacies at the design stage cannot be completely overcome by statistical adjustment procedures during the study or after the end of the study, full consideration of the whole process is required before the start of the study. In this study, we introduce several key items required in design stage. Statistical problems to be considered are discussed with case examples.

Key words : study design, randomization, sample size, statistical analysis