

의학연구 및 의학 논문의 질을 향상시키기 위한 통계적 고려사항

안형진[✉]

고려대학교 의과대학 의학통계학교실

Statistical Consideration for Improving Quality of Medical Study

Hyonggin An, Ph.D.[✉]

Department of Biostatistics, Korea University College of Medicine, Seoul, Korea

서 론

골절학 분야를 포함하여 의과학 전 분야의 연구에서 객관적이고 과학적인 결론을 도출하기 위하여 신뢰성 있는 자료를 수집하고, 처리하고 분석하는 과정은 매우 중요하며 이를 위한 통계학의 올바른 적용은 필수이다. 통계적 방법론은 연구결과의 신뢰성에 큰 영향을 주므로 질 높은 의학연구를 위해서는 연구의 목적에 맞는 연구를 설계하고 수행할 수 있어야 하며 수집된 자료 분석에 적절한 통계적 방법을 선택하고 적용할 수 있어야 한다. 본 중설에서는 의학연구 및 논문작성의 질을 높이기 위한 통계적 고려사항에 관하여 고찰한다. 이를 위하여 연구설계에서의 통계적 고려사항, 통계적 가설검정 및 표본수 계산, 통계적 분석방법의 일반적인 원칙, 의학논문에서 통계방법의 기술 및 결과의 표현, 흔히 범하는 통계적 오류에 관하여 순서대로 살펴본다.

본 론

1. 연구방법(연구설계)

어떤 연구자들은 자료의 질에 관련 없이 복잡한 고급 통계분석 방법을 이용하면 객관적인 연구결과를 도출할 수 있을 것이라고 생각한다. 물론 연구의 목적을 달성하고 과학적인 연구결과를 도출하기 위하여 적절한 통계분석 방법을 선택하여 적용하는 것은 매우 중요하다. 하지만, 이에 못지 않게 중요한 점은 적절한 연구계획과 연구설계를 자료수집 및 분석 전에 세우고 이에 따라 연구를 수행하여 적절한 자료를 수집하는 것이다. 만일 자료가 올바른 과정을 통하여 수집되지 않았다면, 그래서 자료의 질이 낮다면 아무리 복잡한 고급 통계분석 방법을 사용하더라도 그 결과의 신뢰성에는 한계가 있다.

어느 분야든지 연구를 통하여 궁극적으로 밝히고자 하는 것은 원인과 결과의 관계일 것이다. 물론, 인과관계를 밝히기 위해서는 많은 요인들을 고려하여야 하지만 중요한 점 중 하나는 원인과 결과 사이에 발생할 수 있는 편향(bias)을 최소화하는 연구를 수행하는 것이다. 편향이란 실제로 원인이 결과에 주는 영향의 정도와 자료로부터 추정된 영향의 정도 사이의 구조적인 차이를 말한다. 예를 들어, 골절의 유합에 있어 A라는 치료방법의 효과를 보기 위하여

[✉]Address reprint requests to: Hyonggin An, Ph.D.
Department of Biostatistics, Korea University College of Medicine,
73 Incheon-ro, Seongbuk-gu, Seoul 136-705, Korea
Tel: 82-2-2286-1437 · Fax: 82-2-2286-1438
E-mail: hyonggin@korea.ac.kr

Financial support: None. Conflict of interest: None.

A라는 처치방법과 무처치를 비교하는 골절연구를 생각해 보자. 이 연구에서 골절 시점부터 골절 유합 시점까지 걸리는 시간에서의 두 군 간 차이를 검정함으로써 A 처치의 효과를 평가하려고 한다. 실제로 두 군 간의 유합시간의 차이는 7일이라고 하고(물론 이 값은 알 수 없다. 이러한 모집단의 미지의 값은 통계학에서는 모수라고 하고 연구에서는 표본을 이용하여 이 모수를 추정하고자 한다.) 연구에서 추정된 값도 7일이라면 편향이 발생하지 않았다고 한다. 하지만, 만일 추정된 값이 10일이면 편향은 -3일이고 추정된 값이 2일이면 편향은 5일이 된다. 결국, 좋은 연구라는 것은 이러한 편향을 최소화하고 연구의 결론을 도출하는 것이다. 연구의 결과에 편향이 발생하게 되면 실제로 처치의 효과가 없음에도 불구하고 효과가 있다고 결정하거나 거꾸로 실제로 처치의 효과가 있음에도 불구하고 효과가 없다고 연구 결과를 잘못 해석하는 오류를 범하게 될 가능성이 높아진다.

연구에서 이러한 편향이 발생하는 이유는 다양하다. 비교연구에서 편향이 발생하는 가장 주요한 이유는 군 간에 할당 또는 선택의 불균형 때문이다. 예를 들어, 위의 처치 A와 무처치를 비교하는 예제에서 만일 A 처치를 받은 군의 나이가 상대적으로 무처치를 받은 군에 비하여 많고 추후 이러한 사항을 고려하지 않고 두 군의 유합 시간을 비교하면 두 군의 차이뿐만 아니라 결과에 나이의 효과까지 더해져 실제보다 추정된 값이 더 작을 것이며 이는 편향을 야기한다. 이렇듯 결과에 편향을 발생시키는 제3의 변수를 교란변수(confounder)라고 한다. 현재 이러한 교란변수로 발생하는 편향을 최소화할 수 있는 방법으로 무작위배정(randomization)을 이용한 전향적 중재연구가 있다. 이를 무작위배정 통제 시험(randomized controlled trials, RCT)이라고도 하고 근거중심의학에서 가장 높은 레벨의 연구로 분류된다. 하지만, 윤리적인 고려사항, 비용 등 여러 가지 이유로 항상 RCT가 가능한 것은 아니다. 예를 들어, 수술적 처치방법과 약물방법을 비교하는 연구에서는 윤리적인 이유로 무작위 배정이 불가능하다. 이러한 경우에는 중재연구가 아닌 관찰연구(observational study)를 시행할 수 있다. 관찰연구는 연구자가 처치를 중재할 수 없고, 다만 현재 이루어지고 있는 처치와 그 결과를 시간에 따라 관찰할 수밖에 없다. 따라서 처치에서 발생할 수 있는 교란변수의 영향을 연구설계 단계에서 최소화할 수는 없다. 하지만, 연구수행과정 중에 잠재적인 교란변수를 미리 많이 수집할 수 있다면 추후 자료의 분석단계에서 여러 가지 통계적 방법을 이용하여 교란편향을 보정할 수 있다.

연구설계의 종류를 분류하는 방법은 여러 가지가 있을 수 있다. 하지만, 가장 많이 사용하는 분류체계는 처치의 중재여부에 따른 분류인 중재연구와 관찰연구, 연구의 시

점에 따른 분류인 단면연구(cross-sectional study)와 종적연구(longitudinal study)가 있다. 중재연구와 관찰연구를 구분하는 가장 간단한 기준은 연구자가 중재를 시행했느냐의 여부이다. 만일 연구자가 중재를 실시하였으면 중재연구가 되고 그렇지 않고 현상을 관찰만 하였다면 관찰연구가 된다. 임상시험은 의학에서 가장 많이 시행되는 중재연구이다. 중재연구에서 무작위 배정을 실시하게 되면 교란편향을 최소화할 수 있으며 이는 군 간의 비교성을 보장하게 된다. 관찰연구는 연구의 특성상 선택편향을 가지게 되며 이러한 선택편향을 최소화하기 위해서는 짝짓기 방법이나 통계적 보정방법을 이용하여 교란변수를 통제하여야 한다. 이미 언급한 대로 근거중심의학에서는 중재연구가 일반적으로 더 높은 수준의 근거를 창출한다. 하지만 연구결과를 일반 모집단에 적용하는 일반화의 관점에서는 교란변수를 잘 보정할 수만 있다면 관찰연구도 높은 수준의 근거를 창출할 수 있다. 중재연구는 윤리적 이유, 비용, 시간 등의 이유로 연구의 모집단을 잘 대표할 수 있는 표본보다는 선택의 용이성이 높은 표본을 제한적으로 이용할 가능성이 있다. 예를 들면, 무작위 배정 임상시험을 실시하는 경우, 임상시험에서 고려하는 처치는 실제로는 고위험군에서 더 적용되는 경우가 많을 수 있음에도 불구하고 고위험군은 윤리적인 이유로 임상시험에서 제외하는 경우가 일반적이다. 또한, 임상시험의 효율성을 높이기 위하여 환자가 잘 등록될 수 있는 대도시의 대규모 병원에서만 임상시험이 진행되는 경우가 많다. 하지만 임상시험에서 고려하는 처치는 대도시의 대규모 병원에 등록된 환자뿐만 아니라 모든 환자에게 적용하는 것을 목적으로 한다. 만일 대규모의 관찰연구를 모집단을 잘 대표할 수 있는 표본을 대상으로 실시하고 선택편향을 최소화할 수만 있다면 질 높은 연구도 가능하다.

연구의 시점을 기준으로 하는 연구설계의 분류인 단면연구와 종적연구의 구분은 연구의 기간이다. 현상을 한 시점에서 연구하게 되면 이는 단면연구이고 기간을 가지고 연구하면 종적연구이다. 단면연구의 대표적인 예로는 설문조사가 있다. 단면연구는 유병률과 같이 한 시점에서의 의학적 현상을 연구할 때 효율적이나 연구의 특성상 인과관계를 밝히는 데에는 한계가 있다. 종적연구는 다시 전향적 연구(prospective study)와 후향적 연구(retrospective study)로 구분할 수 있다. 현 시점에서 과거시점으로 연구를 진행하거나 과거의 어느 시점에서 현재의 시점까지 연구를 진행하게 되면 후향적 연구이고 현시점에서 연구를 시작하여 미래의 시점까지 연구를 진행하게 되면 전향적 연구이다. 임상시험은 전향적 연구의 대표적인 예이다. 차트리뷰나 사례-대조 연구(case-control study)는 후향적 연구의 대표적인 예이다.

2. 통계적 가설검정 및 표본수 계산

표본수 계산은 연구설계 단계에서 미리 정한다. 연구에 필요한 적절한 표본수를 미리 정하는 것은 매우 중요하다. 좋은 연구계획은 임상적 유의성이 있다면 이를 탐지할 수 있을 정도로(즉, 통계적 유의성을 동시에 보일 정도로) 충분히 큰 규모이어야 하나 임상적 유의성이 없음에도 통계적 유의성을 보일 정도로 너무 큰 규모의 연구는 지양하여야 한다. 위의 골절 유합 처치연구의 예제에서 만일 A 처치가 무처치에 비하여 평균적으로 1일 더 빨리 골절 유합이 된다고 하자. 이러한 차이는 실제로 존재할 수는 있으나 임상적으로는 의미없을 수 있다. 하지만, 만일 표본수가 매우 크다면 이러한 작은 차이도 통계적으로는 유의할 수 있다. 이러한 임상적으로도 의미없는 차이의 통계적 유의성을 보기 위하여 매우 큰 규모의 연구를 진행하는 것은 사회적으로나 과학적으로 아무런 이익을 주지 못하고 많은 수의 연구참여자를 잠재적인 위험에 처하게 할 수 있다. 이는 윤리적인 문제를 야기할 수 있다. 반대로 A 처치가 무처치에 비하여 평균적으로 10일 이상 더 빨리 골절이 유합되면 A 처치는 임상적으로 효과가 있다고 하자. 또한 실제로 A 처치가 임상적인 효과가 있다고 하더라도 만일 표본수가 상대적으로 너무 작다면 연구의 검정력이 너무 낮아 통계적인 유의성을 보이지 못할 수 있다. 따라서, 연구설계 단계에서 믿을 만한 통계적 판단을 가능하게 하는 적절한 표본수를 계산하는 것은 중요하다. 통계적 관점에서 표본수 계산은 검정력 분석과 밀접한 관계를 가지고 있다. 이는 통계적 검정의 개념을 이해하여야 하므로 이 절에서 통계적 검정에 관하여 간략하게 설명하도록 한다.

통계적 가설검정에서 수행하는 연구에서 연구자가 증명하고자 하는 가설을 대립가설(또는 연구가설)이라고 하고 H_A 또는 H_1 으로 표시한다. 만일 이 대립가설을 증명하지 못하고 돌아가야 하는 현재 상태의 가설을 귀무가설이라고 하고 H_0 로 표현한다. 예를 들어, 골절 유합 처치 연구의 예제에서 귀무가설은 'A 처치군과 무처치군 사이에 골절 유합의 평균시간의 차이는 없다(즉 A 처치의 효과는 없다)'이고 대립가설은 'A 처치군과 무처치군 사이에 골절 유합의 평균시간의 차이는 있다(즉 A 처치의 효과는 있다)'이다. 통계적 검정은 귀무가설을 기각하여 대립가설을 지지할 것인지 아니면 귀무가설을 기각하지 못하는지에 대한 의사를 결정하는 과정이다. 통계적 검정(즉 의사 결정)에는 두 가지 오류가 발생하게 된다. 만일 귀무가설이 정말로 사실임에도 불구하고 귀무가설을 기각하는 오류(즉 A 처치의 효과가 없음에도 불구하고 A 처치의 효과가 있다고 결정하는 오류)를 제1종의 오류라고 한다. 제2종의 오류는 만일 귀무가설이 사실이 아님에도 불구하고 귀무가설을 기

각하지 못하는 오류(즉 A 처치의 효과가 있음에도 불구하고 A 처치의 효과가 있다고 결정하지 못하는 오류)이다. 제1종의 오류를 범할 확률을 α 라고 하고 제2종의 오류를 범할 확률을 β 라고 한다.

가설검정에서 가장 좋은 의사 결정은 α 와 β 를 모두 동시에 최소화하는 결정일 것이다. 하지만, α 와 β 를 둘 다 동시에 최소화하는 것은 불가능하다. α 를 작게 만들면 β 가 커지고, β 를 작게 만들면 α 가 커지게 된다. 두 가지 오류를 동시에 최소화할 수 없기 때문에, 통계적 검정에서는 사회적으로 중요하게 생각되는 제1종의 오류를 범할 확률 α 를 연구 시작 전에 미리 작게 결정하고 제2종의 오류를 범할 확률 β 를 최소화하려고 노력하는 접근법을 사용한다. 이렇게 연구시작 전에 미리 결정된 α 를 유의수준이라고 한다(일반적으로 의학연구에서는 0.05를 유의수준으로 이용한다). 다음으로 연구과정에서 수집된 자료와 통계이론을 이용하여 제1종의 오류를 범할 확률을 구한다. 이를 p-값이라고 한다. 다음으로 p-값을 유의수준과 비교하여 더 작으면 귀무가설을 기각하고 유의수준보다 크면 귀무가설을 기각하지 못한다. 사실 통계적 가설검정은 법원에서 재판하는 과정과 매우 유사하다. 법원의 재판과정을 간략히 설명하면 먼저 피고인은 무죄라고 믿는 무죄 추정의 원칙(beyond reasonable doubt)으로부터 시작한다. 그러면 검사가 수집한 여러 가지 증거들을 배심원 또는 판사에게 제시하고 피고인이 유죄임을 증명하려고 노력한다. 이러한 증거를 바탕으로 배심원 또는 판사는 피고인이 유죄인지 아닌지 판단하고 결정을 내리게 된다. 만일 증거가 충분하다고 생각되면 유죄라고 결정을 내리고 증거가 불충분하면 유죄가 아니라고 결정을 내린다. 사실 재판 과정은 피고인의 유죄여부를 결정하는 과정이지 피고인이 유죄가 아니라고 결정을 내렸다고 해서 피고인의 무죄를 증명한 것은 아니다. 이와 비슷하게 통계적 가설검정은 먼저 귀무가설이 사실이라고 믿고 검정을 시작한다(무죄 추정의 원칙과 유사). 이제 연구자(검사)가 올바른 연구과정을 통하여 자료(증거)를 수집한다. 수집된 자료를 바탕으로 통계적 분석을 통하여 p-값이 유의수준보다 작으면 귀무가설을 기각하고(유죄라고 결정을 내리고), p-값이 유의수준보다 크면 귀무가설을 기각하지 못한다(유죄가 아니라고 결정을 내린다). 재판과정과 마찬가지로 통계적 가설검정에서는 귀무가설이 사실이라고 믿고 검정을 시작하였으므로 대립가설을 지지하느냐 아니냐에 관한 결정이지 자료가 대립가설을 지지하지 않는다고 하여 귀무가설을 채택하는 것은 아니다.

이러한 통계적 가설검정의 절차를 요약하면 다음과 같다.

- 1) 통계적 가설검정에 이용할 모수를 정한다(예: 평균, 오즈비, 위험비 등).
- 2) 이 모수를 이용하여 귀무가설과 대립가설을 설정한다(예:

귀무가설-두 군의 평균이 동일하다, 대립가설-두 군의 평균이 동일하지 않다).

- 3) 유의수준 α 를 미리 정한다(일반적으로는 0.05를 사용).
- 4) 통계적 이론을 이용하여 가설 검정에 사용할 검정통계량을 정한다.
- 5) 표본으로부터 검정통계량을 구하고 이에 따른 p-값을 구한다.
- 6) p-값과 α 를 비교하여 α 보다 작으면 귀무가설을 기각하고 그렇지 않으면 귀무가설을 기각하지 못한다.
- 7) 검정결과를 바탕으로 결론을 도출하고 설명한다.

위에서 언급한 바와 같이 가설검정의 과정에는 제1종의 오류만을 이용한다. 하지만 연구자는 또한 유의수준이 주어진 상태에서 제2종의 오류를 범할 확률인 β 를 작게 만들려고 노력해야 한다. 이는 대립가설이 사실인 경우 귀무가설을 기각하는 확률($1-\beta$)인 검정력을 크게 만들려고 노력하는 것과 동일하다. 검정력은 통계적 검정 방법의 종류, 유의수준, 연구의 측정오차, 연구 효과의 크기, 그리고 표본수의 함수로 결정된다. 따라서 만일 이러한 요소가 연구 전에 미리 주어진다면 검정력 함수를 이용하여 연구의 검정력을 계산할 수 있다. 거꾸로 통계적 검정 방법의 종류, 유의수준, 연구의 측정오차, 연구 효과의 크기, 그리고 검정력이 미리 주어진다면 표본수를 계산할 수 있음을 의미한다. 통계적 검정 방법은 연구의 목적, 변수의 종류 등을 고려하여 정할 수 있고, 유의수준은 일반적으로 양측 0.05로 고정하며, 연구 효과의 크기나 연구의 측정오차는 준비연구(pilot study)나 기존의 유사한 연구로부터 추정한다. 검정력은 일반적으로 80% 이상이 될 수 있도록 정한다. 전향적 연구의 경우 중도탈락이 발생할 수 있는데 이러한 중도탈락이 발생하게 되어 미리 정한 표본수보다 분석의 표본수가 작은 경우 미리 정한 검정력을 맞출 수 없으므로 탈락을 고려하여 좀 더 많은 수를 연구에 등록하는 것이 일반적이다. 만일 두 군 비교 연구에서 표본수를 계산하였더니 군당 100명씩(총 200명)의 표본이 필요하다고 계산이 된다면 10%의 탈락률을 고려하는 경우 군당 $100/(1-0.1) \approx 112$ 명(총 224명)을 등록한다.

표본수 계산은 검정력에 근거한 계산법 외에도 추정치의 정도(precision)를 바탕으로 구할 수도 있으나 이 방법은 본 종설에서는 설명하지 않는다. 추정치의 정도를 이용하여 표본수를 계산하는 방법은 표본조사 연구에서 많이 사용한다.

3. 통계 분석의 일반적 원칙

적절한 연구설계에서 자료를 수집하면 연구의 결과를 도출하기 위하여 자료분석을 시행한다. 이때 가장 먼저 시행

해야 할 점은 수집된 자료를 검토하고 기술통계(descriptive statistics)를 이용하여 자료를 요약 정리하여야 한다. 수집된 자료를 검토하는 단계에서는 오류값(errors)을 확인하고 이러한 오류값이 입력오류 등으로 발생한 것인지 원자료와 대조하는 과정을 거친다. 이런 과정이 끝나면 기술통계를 이용하여 자료의 분포적 특성에 관하여 조사한다. 이는 그림이나 요약값을 이용할 수 있다.

일반적으로 연속형 자료의 경우는 히스토그램, 상자수염 그림 등을 이용하여 분포의 형태를 조사할 수 있다. 요약값의 경우는 분포의 위치를 나타내는 평균과 퍼져 있는 정도를 나타내는 표준편차를 이용할 수 있다. 만일 분포가 너무 치우쳐져 있는 경우는 평균은 극단값에 영향을 많이 받기 때문에 중앙값과 사분위수를 이용하여 분포를 요약할 수 있다. 자료가 범주형인 경우는 막대그림을 그려볼 수 있고 절대빈도와 상대빈도를 이용하여 요약값을 제시할 수 있다. 이러한 수집된 자료를 설명하는 요약값은 논문의 표 1로 제시하여 독자들이 자료의 특성을 이해할 수 있도록 돕는다. 이러한 요약값 정리 과정을 통하여 특이값(outlier)을 검출할 수도 있다. 오류값과는 달리 특이값은 극단값이며 잘 발생하지는 않으나 있을 수 있는 값으로 분석에서 임의로 제외하지 말아야 한다. 만일 자료에 특이값이 있는 경우는 일반적으로 포함하여 분석한 결과와 제외하고 분석한 결과를 비교하여 만일 결과에 영향이 없다면 분석의 가정에 좀 더 합당한 결과를 제시한다. 다만 이 경우에도 특이값의 존재와 만일 제외한 결과를 제시하였다면 제외한 이유와 제외하였다는 사실을 논문에 기술하여야 한다.

연구결과의 질은 얼마나 고급스럽고 복잡한 통계방법을 사용하였느냐에 따라 높아지는 것이 아니라 얼마나 적절한 방법을 사용하였느냐에 따라 달라진다. 적절한 통계방법은 연구의 목적, 연구설계방법, 평가변수의 종류, 비교하고자 하는 군의 수, 분석에서 고려하는 변수의 수 등을 종합적으로 고려하여 선택하여야 한다. 만일 골절의 유합에 A 처치의 효과를 검증하기 위하여 A 처치와 무처리군의 유합 시간을 비교하는 관찰연구를 수행하였다고 하자. 유합 시간은 연속형 변수로 두 군 간의 유합 시간의 평균을 비교하는 독립 이표본 t-검정이라는 통계적 방법을 이용하여 군 간 평균의 차이를 비교할 수 있다. 독립 이표본 t-검정의 결과가 유효하기 위해서는 유합 시간의 분포는 정규분포를 따라야 하며 두 군 간의 분산도 동일해야 한다. 하지만, 유합 시간은 일반적으로 정규분포를 따르지 않고 치우친 형태의 분포를 하는 경우가 많고 만일 수집한 자료도 정규분포를 따르지 않는다면 군 간의 분포를 가정하지 않는 비모수적 통계방법인 맨-윌트니 U-검정을 수행하는 것이 적절할 수 있다. 만일 유합 시간이 정규분포를 따르지 않고 또 모든 개체에서 관찰되지 않아 중도탈락(censoring)이

발생하는 경우에는 생존분석을 적용하는 것을 고려할 수 있다. 누적 유효 확률을 카플란-마이어 생존곡선을 이용하여 그림으로 표현할 수 있으며 로그-순위 검정을 이용하여 군 간의 생존곡선을 비교할 수 있다.

또한, 이 예제의 연구는 관찰연구이므로 두 군 간 비교에 교란변수에 의한 편향이 발생할 수 있다. 예를 들어, 두 군 간에 연령, 성별, 골절의 정도, 골절의 종류에 차이가 있고 이러한 제3의 변수들이 결과변수와 관련이 있다면 결과변수의 군 간 비교에 영향을 미치고 따라서 편향을 발생시킬 수 있다. 군 간 비교성을 확보하여 올바른 결론을 도출하기 위해서는 이러한 교란변수를 보정해 주어야 한다. 이러한 교란변수를 보정하는 방법으로 선형회귀분석, 성향점수 방법 등의 통계적 방법론을 이용할 수 있다. 만일 생존분석을 이용하는 경우라면 콕스의 비례위험 회귀분석을 고려할 수 있다. 이렇듯, 적절한 통계분석을 선택하기 위해서는 여러 가지 사항을 고려하여야 한다.

또한, 각각의 통계분석 방법은 가정을 가지고 있다. 예를 들어 독립 이표본 t-검정의 경우 자료는 서로 독립적이어야 하며(독립성), 군 간에 정규분포를 따라야 하며(정규성), 두 군의 분포의 분산은 동일해야 한다(등분산성)는 가정이 필요하다. 만일 분석에 사용된 자료가 이러한 분석 가정을 만족하지 않으면 연구 결과 및 가설검정에 오류가 발생할 수 있다. 따라서, 분석 후에 자료가 분석방법의 가정을 만족하는지 가정 검토를 반드시 수행하여야 하며 가정 검토 결과 자료가 가정을 만족하지 않는 경우는 자료의 변환, 비모수적 방법, 비선형적 방법 등 적절한 대안을 선택하여 재분석을 하여야 한다.

4. 의학논문에서 통계방법의 기술 및 결과의 표현

위에서 언급한 바와 같이 적절한 통계분석을 통하여 자료를 분석하였더라도 만일 작성한 논문에 분석한 통계방법을 올바르게 기술하지 못하거나 결과를 부적절하게 표현하게 되면 애써 연구한 결과의 출판이 거절될 수도 있다. 논문을 작성하는 경우에 올바르게 통계방법 및 결과를 기술하는 방법에 관하여 살펴보자.

연구설계 방법, 자료수집과정, 통계분석 방법은 의학논문의 방법(method) 부분에 자세하게 기술하여야 한다. 수행한 연구가 전향적인지 후향적인지, 중재를 실시하였는지 아니면 관찰연구인지에 관하여 자세하게 기술하고 표본의 선택과정 및 자료의 수집과정에 관해서도 자세하게 기술하여야 한다. 자료의 분석과정은 ‘연구방법’ 섹션 내의 ‘통계분석(statistical analysis)’ 또는 ‘통계방법(statistical methods)’의 섹션에 자세하게 기술한다. ‘통계분석’ 섹션에는 자료분석에 사용한 통계분석 방법, 유의수준, 분석에 사용

한 프로그램 등을 기술하게 되는데, 기술 시 주의할 점은 사용한 모든 통계분석 방법의 이름을 단순히 나열하는 것이 아니라 어떠한 목적을 위하여 어떠한 통계방법을 사용하였는지 자세하게 기술하여야 한다. 예를 들어, 다중선형 회귀 분석을 수행하여 자료를 분석하였다면 어떠한 목적으로 다중선형회귀 분석을 수행하였는지 기술한다. 일반적으로 다중선형회귀 분석은 여러 가지 목적을 위하여 분석에 사용할 수 있다. 크게 원인과 결과의 영향을 평가할 때 교란요인을 통제하거나, 결과와 관련된 중요한 위험요인을 파악하거나, 다양한 요인을 이용하여 결과에 대한 예측식을 구축할 때 다중선형회귀 분석을 사용한다. 따라서 이러한 다양한 목적 중에 어떤 특정한 목적을 위하여 사용하였는지 기술해 주어야 한다. 또한 어떠한 종속변수와 독립변수들이 다변량 회귀분석에서 고려되었는지 기술하여야 한다. 일반적으로 많이 사용되고 독자들에게 친숙한 통계적 방법에 관해서는 방법 자체를 설명할 필요는 없지만 만일 분석에 사용한 방법이 일반적이지 않고 독자들에게 친숙하지 않다고 판단되면 사용한 분석방법 자체에 관하여 참고 문헌과 함께 간략하게 설명해 주는 것이 독자들의 이해를 돕는다. 통계분석 섹션을 기술하는 가장 좋은 전략은 만일 독자들이 분석에 사용한 원자료가 있다면 기술된 통계방법을 시행하여 결과를 재현할 수 있도록 명확히 기술하는 것이다.

‘결과(result)’ 섹션은 분석의 결과를 표나 그림을 이용하여 제시하고 설명한다. 표나 그림을 작성하는 원칙은 독자가 본문을 읽지 않아도 표나 그림만 보고 분석결과를 충분히 이해할 수 있도록 명확하게 제시하는 것이다. 이를 위해서는 주석을 적극 활용하여 표나 그림을 설명한다. 숫자를 제시하는 경우 소수점은 자료의 단위와 임상적으로 관련 있는 정도만 제시한다. 일반적으로는 소수점 두 자리이면 충분하고 아무리 길어도 세 자리를 초과하지 않는 것이 좋다. 그림의 경우는 각 축에 해당하는 변수명과 그 단위를 분명히 제시하며 범례는 구체적으로 제시한다. 결과 섹션의 본문에는 표나 그림에 제시된 숫자를 다시 반복하는 것이 아니라 가능하면 표에 제시된 값들의 질적인 표현에 중점을 둔다. 통계적 검정을 한 경우에는 p-값을 제시한다. p-값을 제시하는 경우 주의할 점이 몇 가지 있다. 의학연구자들이 많이 사용하는 통계패키지인 SPSS는 매우 작은 p-값을 0.00으로 제시하는데 이를 표나 그림에 그대로 옮기지 말고 부등호를 이용하여 <0.001 등으로 표현한다. 또한, <0.05나 <0.01 등의 구간을 이용하는 것보다는 가능하면 p-값은 원 값을 그대로 제시한다. 간혹 유의한 p-값만 제시하고 유의하지 않은 경우에는 ‘NS (not significant)’ 등으로 표현하는 경우가 있는데 유의하지 않은 p-값도 원 값을 제시한다. 또한, p-값과 더불어 모수의 추정치(예: 오즈비)

를 제시하고 95% 신뢰구간도 같이 제시하여 준다. 신뢰구간은 연구의 크기를 보여주는 매우 좋은 값들이다. 같은 p-값이더라도 좀 더 좁은 신뢰구간을 보이는 연구는 그 연구의 크기가 상대적으로 크다.

‘고찰’ 섹션에는 연구의 독창성, 연구결과의 강점 및 시사점에 관하여 충분히 기술한다. 또한 만일 있다면 연구설계, 표본수, 일반화 등에 있어서의 제약을 설명하고 이러한 제약이 연구의 결과에 주는 영향에 관하여 기술하여 준다.

5. 흔히 범하는 통계적 오류

이 장에서는 연구자들이 흔하게 범하는 통계적 오류에 관하여 간략하게 설명한다. 의학연구에서 흔히 범하는 통계적 오류 중 하나는 연구목적과 그에 따른 결과변수가 불분명한 경우이다. 결과변수는 연구와 명백하게 관련이 있어야 하며 측정에 오류가 없어야 한다. 단지 유의성이 높을 수 있는 결과변수의 선택은 지양하여야 한다. 또한 분석에 사용한 표본수를 명확하게 제시하여야 한다. 어떤 경우, 결측값으로 인하여 실제 수집한 표본의 수보다 분석에서 사용한 표본의 수가 작을 수 있는데 이때 이러한 표본수를 명확하게 제시하여야 한다. 또한 결측이 많은 경우에는 결측으로 인하여 결과에 미칠 수 있는 영향에 관해서도 면밀하게 분석하고 결측으로 인하여 편향이 발생할 수 있는 경우에는 결측을 고려한 분석을 시도한다. 자료에 이상값이 있는 경우, 이 값을 제외하고 분석하는 경우에도 제외한 이유를 포함하여 제외하였다는 사실을 기술하여야 한다. 전향적인 연구에서는 가능하면 표본수 계산에 관해서도 기술한다.

무작위 배정을 실시한 경우에는 무작위 배정 과정을 상세히 설명하여야 하며 무작위 배정 후 연구과정에서 중도 탈락(drop-out), 처치 비순응(non-compliance)과 같은 무작위 배정의 손상이 발생한 경우 이에 대한 설명도 기술한다.

연구과정에서 부집단 분석을 실시하는 경우도 있는데 계획되지 않은 부집단 분석은 제1종의 오류를 증가시키게 되어

실제로는 유의하지도 않은데 유의하다고 판단할 가능성을 높인다. 따라서 계획서에 미리 계획되지 않은 부집단 분석을 자료수집 후에 자료를 보고 실시하는 것(이를 post-hoc subgroup analysis라고 함)은 지양한다.

또 먼저 작은 표본을 이용하여 자료분석을 실시하고 만일 유의한 결과를 얻지 못하는 경우 사후에 자료를 추가하여 분석하는 경우가 있는데, 이 또한 제1종의 오류를 증가시키며 연구자에게 유리한 자료를 추가할 가능성이 높으므로 연구 윤리 차원에서 문제가 발생할 수도 있으므로 사후 자료추가는 하지 않는다.

잘못된 통계분석을 사용하는 경우도 빈번하다. 위에서 설명한대로 통계분석의 일반적 원칙을 고려하여 적절한 통계분석을 선택하여 자료를 분석하고 분석 후 자료가 분석의 가정을 만족하는지 꼭 검토한다.

이차자료 또는 표본조사 자료를 이용하여 연구를 진행하는 경우, 자료를 제공할 때 가중치를 같이 제공하는 경우도 있다(예, 국민건강영양조사). 이런 경우에는 가중치를 고려한 분석을 실시하여야 편향이 최소화된 결과를 도출할 수 있다. 관찰연구를 포함하여 무작위 배정이 되지 않은 연구에서는 분석에서 반드시 잠재적인 교란변수를 고려하여 분석해야 선택편향을 최소화할 수 있다.

요 약

본 종설에서는 의학연구 및 의학논문의 질을 향상시키기 위한 통계적 고려사항에 관하여 간략하게 살펴보았다. 좀 더 질 높은 의학 연구를 수행하기 위해서는 연구설계 단계에서 연구의 목적을 명확하게 설정하고 이에 알맞은 연구설계를 계획하여 자료를 수집하고 적절하고 올바른 통계분석을 통하여 과학적인 결론을 도출하고 해석해야 한다. 또한 질 높은 논문을 작성하기 위해서는 이러한 일련의 과정을 명확하게 논문에 기술하여야 한다. 본 종설이 골절학 분야를 포함한 의학연구 수행 및 논문 작성에 도움이 되길 바란다.