

## 논문심사와 통계

가톨릭대학교 의과대학 의학통계학교실

박 용 규

### 요 약

논문을 심사하는 사독위원의 입장에서 통계적인 부분의 평가를 어떻게 할 것인지에 대해, 비교적 중요하게 생각되는 점들을 간략히 소개한다. 이 내용만을 이용하여, 논문의 통계적 오류를 완벽하게 찾아내기를 바라는 것은 지나친 요구가 되겠지만, 보다 전문적인 내용은 학회의 통계자문위원에게 문의한다면, 심사시 애로점의 상당 부분은 해결되리라 믿는다.

또한 각 분석방법을 구체적으로 설명하지 않았기 때문에, 실제 논문을 작성하는 연구자의 입장에서는 그다지 도움이 되지 않는겠지만, 논문을 완성한 후 투고하기 전에 참고한다면, 심사과정에서 지적될 수 있는 사항들을 미리 파악할 수 있어, 그만큼 시간도 절약할 수 있으리라 생각된다.

### 서 론

사독위원이 가정의학회지에 투고된 논문을 심사할 때에는, 그 논문의 의학적 측면에서의 중요성과 함께, 결론에 이르기까지의 과정이 적절한가를 살펴 보게 된다. 그리고 의학을 전공한 대부분의 사독위원들은 논문이 갖는 의학적인 가치를 평가하는 것보다 연구방법, 특히 연구의 설계, 자료수집 과정, 통계분석방법 등이 합리적인지를 평가하는데 더 어려움을 느낀다. 그렇지만 논문의 게재여부를 결정하여야 할 입장에서는, 자신의 전공이 아니라 해서 논문에서 사용된 통계적인 방법에 대한 평가를 그냥 지나칠 수도 없고, 그렇다고 통계학 공부를 다시 시작할 수도 없는 노릇이다.

논문의 통계적 문제점을 심사하는 가장 손쉬운 방법은 의심스러운 부분에 대해 학회의 통계자문위원을 활용하는 것이다. 현재 많은 분들이 통계문의를 하고는 있지만, 대부분의 사독위원들은 심사에 소요되는 시간이 길어지고 번거로우며, 통계자문위원과의 개인적 친소관계를 고려해 주저하는 것으로 보인다.

다. 이 글의 목적은 그동안 학회지의 통계논평을 진행하면서 비교적 자주 언급했고, 중요하다고 생각되는 몇가지 통계적 문제점들을 간략히 소개하여, 심사에 도움을 주고자하는데 있다.

### 본 론

#### 1. 통계적 기법이 사용된 논문은 어떻게 찾아가?

논문의 통계부분을 심사할 때에는 그 논문에 통계적인 내용이 있는지를 먼저 파악해야 한다. 이를 위해서는 다음 내용을 살펴보는 것이 도움이 된다.

- 요약문 또는 본문의 결과부분에 P값이 제시되어 있는지 본다. P값이란 저자가 연구한 내용 중 의미있는 발견이 있었다는 증거를 통계적 검정을 통해 제시한 것이다.

- 요약문의 결론이나 본문의 고찰부분에 “유의한 차이” 또는 “유의한 관련성”이라는 표현이 있는지 본다. 여기서 “유의한(significant)”이란 “통계적으로 유의한(statistically significant)”을 줄여서 표현한 것으로, 이 역시 통계적 검정을 했다는 뜻이 된다.

- 표나 그림에 주석기호(\*, †, ‡, §, ||, ¶, \*\* 등)와 그 기호에 대한 설명에 P값이 있는지 살펴본다.
- 본문의 방법부분에 통계 검정법의 이름이 있는지 살펴본다. 통계 검정법의 이름을 잘 모른다하더라도, 대부분 “~ 검정”, “~ 분석”이라는 표현으로 되어 있으므로 쉽게 구분할 수 있다.
- 본문의 방법부분에 “랜덤추출(random sampling)”, “표본” 등의 표현이 있는지 살펴본다.

## 2. 굳이 통계적 기법을 사용할 필요가 있었는가?

위와 같은 절차를 거쳐 저자가 통계적 방법을 사용한 것으로 파악된다면, 그 논문에서 통계적인 분석이 과연 필요한 것이었는지를 다음과 같이 살펴본다.

- 지역적으로 또는 시간적으로 지나치게 제한된 연구대상을 이용하지 않았는지, 특히 처음부터 관심을 가졌던 연구대상자들을 한 사람도 빠뜨리지 않고 조사한 것이 아닌지 살펴본다. 만약 연구자가 고려했던 대상자가 모두 연구에 참여했다면, 표본이 아닌 모집단을 조사한 것이므로, 통계적 검정(P값의 제시)이 필요없으며 단순히 집계한 표만 제시하면 된다.<sup>1)</sup>
- 연구대상자들을 표본으로 간주할 수 있다고 하더라도, 그 표본을 뽑는 방법에 “편의적 추출”이라는 표현이 있는지 살펴본다. 모든 통계적 검정의 결과를 나타내는 P값이란 확률값(probability value)을 말하며, 이는 “랜덤추출”, 즉 실제 연구대상이 된 사람들이 치우침(bias)없이 뽑혔다는 가정을 근거로 하여 계산된 값이므로, 편의적으로 추출된 표본에 대해서는 통계적 검정을 하지 않아야 한다. 또한 편의적 추출에 의한 결과는 모집단을 제대로 반영하지 못한다는 점에서 문제가 많다고 할 수 있다.<sup>2)</sup>
- 광범위한 지역을 대상으로 어떤 질병의 유병률을 알아내기 위해, 그 지역 주민 모두를 조사했다면, 모집단을 조사(전수조사)한 것이므로 통계적 검정을 하지 않는다. 이때 우편이나 전화를 이용하여 조사한다면, 계획했던 모든 대상자들로부터 응답을 얻지 못하는 경우가 많다. 이를 표본으로 간주하여 통계적 검정을 하는 경향이 있으나, 엄격하게 평가한다면 랜덤추출한 표본이 아니므로 잘못되었다고 할 수 있다. 이 경우에는 무응답들의 특성과 응답자들의 특성을

비교하여, 수집된 자료가 충분한 대표성을 가진다는 점을 간접적으로나마 증명해야 한다.<sup>3)</sup>

- 위의 연구를 전수조사 대신, 처음부터 각 지역의 인구 또는 특성의 분포에 맞춰 비례적으로 표본을 뽑아 조사(표본조사, sample survey)했다면, 통계적 검정결과를 별도로 제시할 필요가 없다. 이 때에는 추출하고자한 표본의 수가 어떤 근거에서 산출된 것인지, 즉 표본오차(sampling error)를 제시하는 것만으로 충분하다.<sup>3)</sup>

· 이와는 반대로, 연구대상이 표본인데도 불구하고 통계적 검정을 하지 않았다면, 특별한 경우를 제외하고는 이의를 제기해야한다. 이는 연구자가 내린 결론이 증명되지 않았기 때문이다. 여기서 특별한 경우란, 매우 작은 수의 대상자들로 연구가 이루어져, 통계적 검정방법으로 그 의의를 찾는다는 것이 무의미하지만 (실제로 표본수가 작을 때에는 통계적으로 유의한 결과를 얻기 힘들다), 연구를 시도한 것 자체만으로도 의학적으로 충분한 가치가 있다고 판단되는, 중간보고서, 증례보고, brief communication 등을 말한다.<sup>4)</sup>

## 3. 통계적 기법에 대한 평가는 구체적으로 어떻게 할 것인가?

이제는 통계적 분석이 반드시 필요했던 논문을 대상으로, 사용된 통계적 기법이 적절한 것인지 검토해보자. 여기서 말하는 통계적 기법이란 특정 통계검정법이나 분석법만을 언급하는 것이 아니라, 연구방법이나 결과부분에 제시된 내용 중 통계와 관련된 모든 내용을 포함한다. 이 부분은 특히 통계적인 지식을 많이 필요로 한다.

1) 부정적 발견<sup>5)</sup>: 연구자의 주장의 독창성 또는 논문으로서의 가치를, 특정 집단들간에 반응값들이 통계적으로 유의한 차이가 있거나 연구자가 고려한 몇가지 요인들간에 유의한 관련성이 있다는 결론을 이끌어내었는가로 판단하는 경향이 있다. 물론 독창성을 과거의 이론 또는 기존의 지식과의 차별로 볼 때, 대부분의 연구내용은 의미있는 차이를 발견하는데 그 목적이 있겠지만, 만약 “서로 차이가 있다” 또는 “관련이 있다”는 것이 과거의 이론이나 지식이라면, 이를 부정하는, 즉 “서로 차이가 없다” 또는 “관

련성이 없다”는 주장이 오히려 독창성을 갖게 될 것이다. 다시 말해, 부정적 발견(negative finding)만으로 구성된 논문도 나름대로 가치는 있다고 본다.

그러나 부정적 발견을 주장한 논문들은 그 주장의 근거를 제시하지 않은 경우가 많이 있다. 여기서 부정적 발견에 대한 근거란, 제 2종의 과오를 범할 확률( $\beta$ ) 또는 검정력(power,  $1 - \beta$ )을 말한다. 이 값들은 부정적 발견을 주장할 때 그 주장을 신뢰할 수 있는 정도를 나타내는 것으로, 흔히 긍정적 발견(positive finding)을 주장할 때 제시하는 유의수준과 같은 성격을 갖는다.

2) 연구대상자들의 선택: 연구방법의 첫부분에 기술되어 있는 실제 조사대상자들의 성격이, 논문의 제목이나 요약문의 연구배경 그리고 본문의 고찰부분에서 저자가 주장하는 대상과 일치하는지 검토해야 한다. 즉, 목표모집단(target population)과 표본모집단(sample population)이 일치해야 한다.<sup>6)</sup>

연구계획이나 집단의 구성이, 연구에서 밝히고자 하는 주요 요인 이외에는 그 성격이 서로 동질적이 되도록 노력했는지 검토한다. 전향적인 연구(prospective study)에서는 실험계획(experimental design)을 통해서, 후향적인 연구(retrospective study)에서는 짝짓기, 블록화, 층화 등을 통해 비교집단간의 동질성을 유지할 수 있다. 만약 이러한 동질성을 확보하지 못했다면 다변량 분석과 같은 통계적 기법으로 집단간의 이질성을 보정(adjust)해야 한다.

3) 분석방법의 서술<sup>6)</sup>: 주로 연구방법의 끝부분에 기술되어 있으며, 표나 그림, 그리고 결과에 서술된 내용을 읽기 전에는 그 방법의 적절성을 파악하기 어렵다. 따라서 분석을 하기 위해 측정값들을 변환했는지, 특정 cut-off 값에 따라 집단을 구분했는지 등에 관해 구체적으로 설명되어 있는지 우선 살펴본다.

다음에는 서술된 통계검정방법의 이름이 정확한지 검토한다. 만약 그 검정법에 대해 잘 모른다면 한국통계학회가 편찬한 통계용어사전을 참고할 것을 권한다. 가능하다면 기술된 검정방법들이 결과부분에서 실제 사용되었는지 비교해 본다.

긍정적 발견을 위한 논문은 유의수준이, 부정적 발견을 위한 논문은 검정력이 각각 제시되었는지, 그리고 여기서 제시한 유의수준이 결과부분에서 일관성

있게 사용되었는지 살펴본다. 예를 들어, 유의수준을 5%로 한다고 기술한 후, 실제 결과에서는 10% 수준에서 유의하다라는 결론을 내리는 경우도 가끔 발견된다.<sup>7)</sup>

4) 검정과 결과제시방법의 적절성: 논문을 직접 작성한 저자보다 통계학에 관한 지식이 더 많아야 올바른 지적을 할 수 있겠지만, 여기서는 통계적 문제점을 비교적 쉽게 판단할 수 있는 방법을 몇가지 소개하기로 한다.

(1) 변수의 척도(분포)와 검정법<sup>6)</sup>: 혈압, 키, 몸무게와 같은 측정값들은 연속적이고, 정규분포를 한다고 가정하고 모수적인 방법을 사용한다. 여기서 모수적인 방법이란 t 검정, 분산분석, 회귀분석, 상관분석 등을 말한다. 연속적인 값이지만 정규분포를 하지 않는 측정값에 대해서는 비모수적인 방법을 사용한다. 심사하는 입장에서는 원자료를 직접 볼 수 없으므로, 표나 그림에 제시된 값의 표준편차가 평균보다 지나치게 크거나, 좌우 비대칭인 분포를 하면 정규분포가 아니라고 판단하면 된다. 비모수적인 방법에는 Wilcoxon의 순위합 검정, 부호순위 검정, Kruskal-Wallis 검정, Spearman 상관계수 등이 있다.

모수적인 방법이 사용된 경우 표에 제시된 값들은 주로 평균과 표준편차가 되며, 비모수적인 방법에는 중앙값과 범위가 주로 제시된다.

범주형 자료는 빈도수와 비율로 제시되며, 카이제곱 검정이 주로 사용되고, 표본수가 작은 경우에는 Fisher의 exact test를 사용한다.

(2) 비교집단의 수와 검정법<sup>6)</sup>: 비교집단의 수가 둘이면 t 검정이나 Wilcoxon 검정이, 세 집단이상이면 분산분석이나 Kruskal-Wallis 검정 또는 Friedman 검정이 사용된다.

세 집단이상이 있을 때 각 두 집단간의 구체적인 비교는 다중비교법(multiple comparison)을 사용하여 P값을 조정해야 한다. Scheffe, Duncan, Tukey, Dunnett, Bonferroni의 방법이 흔히 사용된다.

(3) 관련된 자료의 검정법<sup>6)</sup>: 동일한 사람을 대상으로 반응값을 여러 번 관찰했거나, 짝을 지어 연구한 경우에는 관련된 자료의 분석방법을 사용해야 한다. 이러한 방법으로는 paired t 검정, Wilcoxon의 부호순위 검정, 랜덤화 블록계획법, Friedman 검정, 반

복측정 분산분석법 등이 있다.

(4) 회귀분석: 표에 제시된 값이 표준화 회귀계수가 아닌, 원자료의 척도를 사용한 회귀계수인지 살펴 보아야 한다. 특히 통계분석에 사용한 프로그램이 SPSS인 경우에는 표준화 회귀계수를 제시한 경우가 많이 있다. 표준화 회귀계수는 절대값이 대부분 0에서 3사이의 작은 값들을 가지므로 쉽게 파악할 수 있다.<sup>7)</sup>

동일한 내용에 대해 상관분석과 회귀분석을 모두 사용하였다면, 그 연구자는 분석목표와 방향을 확실히 정하지 않았다고 생각해야 한다. 상관분석은 상호 대등한 두 변수간의 관련성을, 회귀분석은 반응변수와 그것을 설명하는 변수를 구분하기 때문이다.<sup>6)</sup>

회귀분석의 결과를 설명할 때는 회귀계수들의 통계적 유의성과 반응변수의 변이를 설명할 수 있는 정도를 나타내는 결정계수( $R^2$ )를 많이 이용한다. 여기서 결정계수는 변수들간의 실제적인 관련성 정도를 주로 반영하고, 회귀계수의 유의성은 표본수에도 많은 영향을 받는다. 따라서 표본수가 큰 경우에는 회귀계수의 유의성에만 관심을 두지 말고, 결정계수가 충분히 큰지 살펴야 한다.<sup>7)</sup>

다중회귀분석결과를 나타내는 표에 편결정계수(partial  $R^2$ )를 제시하기도 하지만, 각 변수가 반응변수를 설명하는 정도는  $R^2$ 를 이용하여 서술해야 한다.

## 결 론

논문을 심사하는 사독위원의 입장에서 통계적일 부분의 평가를 어떻게 할 것인지에 대해, 비교적 중요하게 생각되는 점들을 간략히 정리해 보았다. 물론 이 글에 언급된 내용만을 이용하여, 논문의 통계적 오류를 완벽하게 찾아내기를 바라는 것은 지나친 요구가 되겠지만, 본문 중에 언급한 것처럼 학회의 통

계자문위원에게 문의하는 보완적인 방법을 취한다면, 심사시 애로점의 상당 부분은 해결되리라 믿는다. 또한 각 분석방법을 구체적으로 설명하지 않았기 때문에, 실제 논문을 작성하는 연구자의 입장에서는 그다지 도움이 되지는 않겠지만, 논문을 완성한 후 투고하기 전에 참고한다면, 심사과정에서 지적될 수 있는 사항들을 미리 파악할 수 있어, 그만큼 시간도 절약할 수 있으리라 생각된다.

이 글에서 언급하지 않은 내용 중, 표와 그림의 작성법에 관한 것은 학회지의 투고규정에 상세히 제시되어 있어 생략하였으나, 표에 제시된 숫자들의 합이 표본수와 동일한지, 백분율의 합이 100이 되는지는 반드시 확인할 필요가 있다. 또한 표와 그림이 동일한 내용을 전달하고 있다면 분명히 중복표현이므로 수정을 요구해야 할 것이다. 그 외에 통계적 오류에 관한 보다 구체적인 내용은 학회지의 논평을 참고로 하기 바란다.

## 참 고 문 헌

1. 박용규. 2001년 11월 게재논문의 통계적 오류. 가정의학회지 2001;22(12):1838-9.
2. 박용규. 2002년 4월 게재논문의 통계적 오류. 가정의학회지 2002;23(5):657-9.
3. 박용규. 2002년 2월 게재논문의 통계적 오류. 가정의학회지 2002;23(3): 381-3.
4. 박용규. 2002년 5월 게재논문의 통계적 오류. 가정의학회지 2002;23(6):804-6.
5. 박용규. 2002년 1월 게재논문의 통계적 오류. 가정의학회지 2002;23(2):242-4.
6. 대한가정의학회 간행위원회. 사독위원 편람. 서울: (주) 메드랑; 2002.
7. 박용규. 2002년 3월 게재논문의 통계적 오류. 가정의학회지 2002;23(4):531-3.