

■ 논 평

## 2002년 3월 게재 논문의 통계적 오류

### 박 용 규

가톨릭대학교 의과대학 의학통계학교실

#### 서 론

2002년 3월 가정의학회지에 게재된 총 8편의 원자들 중, 이덕진 등의 “임상적 혈청학적으로 확진된 홍역환자의 사례”는 사례연구에 해당되는 논문이었고, 이인구 등의 “가정폭력에 대한 전공의의 태도와 지식”과 최지호 등의 “한국형 가족기능 평가도구의 신뢰도와 타당도”에서는 통계분석방법상의 특별한 문제점이 발견되지 않았다. 나머지 5편의 논문에서는 다음과 같은 통계적 오류가 발견되었다.

- 회귀분석 결과의 제시방법
- 유의수준
- 통계적 유의성과 임상적 중요성
- 통계적 표현의 기술방법
- 로지스틱 회귀분석에서 0의 처리

다음은 각 논문에 나타난 통계적인 문제점들을 제시한 것이다.

#### 각 논문에서의 통계적 오류

##### 1. 손중천 등, C-반응성단백과 심혈관질환 위험인자와의 관계

1) 표 3의 다중회귀분석 결과에서 표준화 회귀계수(standardized coefficient)는, 다중회귀분석에 사용된 여러 독립변수들의 반응값에 미치는 영향력을 상대적

으로 비교할 수 있는 값이지만, 원래 변수들이 가진 척도를 알 수 없기 때문에 회귀식을 구할 수 없고, 따라서 예측(prediction)에 사용될 수 없다는 단점이 있다.<sup>1)</sup>

2) 표 3에 있는 값 중에서 수축기 혈압의 표준화 회귀계수의 값 0.045는 잘못 기입된 것으로 보인다. 본문의 해당 내용과 표 3의 t 값 (1.656)으로 미루어 볼 때, 0.04보다는 작은 값이 되어야 한다. 표준화 회귀계수와 검정통계량 t값의 크기는 서로 비례하기 때문이다.

3) 다중회귀분석을 한 결과  $R^2=0.04$  (4%의 설명력)를 얻었다. 이처럼 설명력이 낮아도 불구하고, CRP의 변화를 유의하게 설명해 주는 변수가 네 가지 (체질량지수, 총콜레스테롤, 백혈구, 고밀도지단백질)나 된다는 것은, 총 연구대상자의 수가 3,425명으로 매우 많기 때문으로 보인다. 이러한 현상은 표 2의 단변량분석에 제시된 값들이 각 요인별로 큰 차이를 보이고 있지 않지만, p값은 대부분 유의하게 나타난 것으로도 알 수 있다.

4) 이처럼 통계적 검정은 아무리 작은 차이라도 표본수만 크다면 유의한 결과를 가져온다. 의학인이 아닌 입장에서 표 2에 나타난 정도의 차이가 과연 어떤 의미를 가질런지는 알 수 없지만, 연구자가 임상적으로 의미있다고 생각하는 차이를 미리 설정한 후, 표본이 그만큼의 차이를 보이는지를 검정하는 방법을 고려해 볼 것을 권한다.<sup>2)</sup>

2 송찬희 등, 종합병원에 근무하는 여성근로자의 체력요인과 피로도와의 관련성

1) 자료분석방법과 결과에 있는 “교호작용을 배제한 ... 관련성을 조사하기 위하여 동시투입법 다중회귀분석을 시행 ...”라는 표현 중 “교호작용을 배제한”은 “상호관련성을 고려하여”로, “동시투입법”은 삭제하는 것이 좋다.

2) 실제 다중회귀분석을 하는 과정에서 독립변수들 간의 교호작용을 나타내는 변수를 새롭게 정의한 후, 그 변수를 분석에 사용했다면 위의 표현은 옳은 것이지만, 단순히 여러 변수들을 동시에 회귀분석에 사용했다면, 그 변수들은 혼동변수(confounding variable, 또는 교란변수)로 처리된 것이다.

3) “동시투입법”은 회귀분석을 할 때 여러 변수들을 동시에 사용했다는 뜻으로 보이지만, “다중회귀분석”의 “다중”이라는 용어 자체가 이미 그 뜻을 포함하고 있다.

3 심재윤 등, 금연클리닉에서 금연성공과 관련된 요인

1) 통계분석방법에서 “분포의 비교에는 카이제곱검정 ...”은 “비율 또는 빈도의 비교에는 카이제곱검정 ...”로 하는 것이 보다 정확한 표현이다.

2) 유의수준을 0.05로 한다고 기술한 경우, 이후에 제시되는 모든 검정의 유의성은 0.05를 기준으로 판단되어야 한다. 표 4에서 흡연관련질환의 유무는 p값이 0.068인데도 제시되어있다. 이는 전진선택법(forward selection)에 의한 다중 로지스틱 회귀분석에서의 선택기준이라고 한 0.05에서 벗어나며, 표 아래의 설명 “유의한 결과만 제시”와도 일치되지 않는다.

3) “다변량분석을 위해, 분산분석표의 F값에 가장 큰 기여를 하는 ... 로지스틱 회귀분석을 시행 ...”에서 “분산분석표의 F값”은 회귀분석에서 사용된다. 로지스틱 회귀분석에서는 “카이제곱 값”이 사용된다.<sup>3)</sup>

4 김형준 등, 청소년들의 인터넷 사용실태 및 중독관련 요인 분석

1) 표 6의 백분율은 전체 addict 또는 non-addict

의 수에 대한 요인별 비율을 계산한 것이다. 그러나 본문에서 표 6을 서술한 부분에서는 표에 제시된 백분율을 사용하지 않았을 뿐 아니라, 기술방법도 각 요인에 따라 다르게 되어있다.

2) 먼저 본문에서의 성별, 나이별 백분율은 각 요인별 대상자수에 대한 addict의 비율이며 (예를 들어, addict수/남자수), 이용횟수나 사용장소는 표 6과 같이 (요인별 분포/addict의 수)를 언급하였다.

3) 표 6과 해당 본문의 서술은 본문의 성별, 나이별 백분율과 같이, 각 요인별 addict의 비율로 제시하는 것이 더 좋았을 것이다.

4) 표 8의 로지스틱 회귀분석 결과 중 web-site의 sex-related site와 place의 Ect.의 odds ratio와 95% CI 값이 제대로 추정되지 않은 이유는, 해당 분포에 0(zero)이 있기 때문이다. (표 6과 7 참조) 이와 같이 0이 있는 경우에는 각 빈도에 0.5를 모두 더한 후 추정값을 구하는 것이 일반적인 방법이다.<sup>4)</sup>

5 박일환 등, 금연 프로그램에서 파트너 지지가 금연 효과에 미치는 영향에 대한 메타분석

1) 본 학회지에 처음으로 게재되는 메타분석에 관한 논문으로 생각된다. 많은 연구자들이 메타분석이라는 용어에는 어느 정도 익숙하지만, 실제로 분석하는 방법에는 아직 생소하다는 점에서, 메타분석의 방법이 보다 구체적으로 기술되었으면 하는 아쉬움이 있다.

2) 예를 들어, 본 연구에서 사용된 합병(통합) 교차곱비의 계산에 peto의 방법, Mantel-Haenszel의 방법 또는 DerSimonian과 Laird의 방법 중 어느 것이 사용되었는지 등의 보다 명확한 방법이 기술되었으면 한다.<sup>5)</sup>

3) 본문에서 교차비는 교차곱비 또는 승산비로 표현하고, 결과 서술 중 6-9개월째 교차곱비 1.08과 95% 신뢰구간 0.81-1.44는, 그림 1의 Total에 제시된 값과 서로 일치하지 않는다.

결론

위에서 언급한 메타분석(meta analysis)<sup>5)</sup>에 대해 간단히 소개하는 것으로 결론을 대신한다. 메타분석

은 ‘meta’가 뜻하고 있듯이 (meta는 after의 뜻) 특정분야 또는 주제에 관하여 지금까지 누적된 연구결과들을 종합할 목적으로, 독립적으로 이루어진 개별 연구의 결과들을 한데 모아 통계적 종합분석을 시도하는 ‘분석의 분석’이다. 다시 말해, 동일한 주제에 대해 그 동안 수행되었던 많은 연구결과들이 서로 상이한 결론들을 제시하고 있다면, 도대체 어떤 주장이 옳은 것인지 확인해 보려는 의도에서, 또는 과거 연구들이 비슷한 결론을 내리고 있을 경우에는 관심있는 요인의 효과를 구체적인 하나의 수치로 제시하고자 하는 의도에서, 연구자 자신이 직접 실험이나 조사를 하는 대신, 이미 발표된 다른 연구결과에 제시된 자료, 통계량, p값 등을 이용하여 하나의 종합적인 결론을 내리고자하는 방법이다.

메타분석 방법으로 연구결과들을 종합하면, 개별 연구자의 주관이나 편견을 상당부분 줄일 수 있고, 후속 연구자가 다시 반복하여 검토, 확인할 수 있게 해 준다. 메타분석의 장점으로는 첫째, 선행연구의 개관 혹은 종합이 좀 더 체계적으로 이루어질 수 있다. 즉, 작은 표본수, 한정된 연구계획방법 등의 제약조건으로 인해 다소 미약한 결론을 도출할 수밖에 없는 각 연구들을 종합함으로써, 특히 치우쳐지지 않은 연구대상자들을 선정할 수 있어 보다 객관적인 추론을 가능하게 한다. 둘째, 동일한 주제에 대한 개별 연구를 종합한다는 것은 표본의 크기가 늘어난다는 점에서 통계적 검정력(power)을 높일 수 있다.

이러한 메타분석에 대한 비판으로는, 서로 비교할 수 없는 다른 성질의 연구결과들을 종합하려고 할 수

있고, 종합하려는 논문들의 우수성(질)이 서로 차이가 있을 수 있음에도 불구하고 아무런 구분없이 사용하려고 할 수 있으며, 문헌수집이 완벽하지 않아 정작 중요한 연구논문들이 누락될 수도 있다는데 있다. 특히, 각 저널이나 연구자들이 부정적 발견(negative finding)의 논문을 기피하려는 경향이 어느 정도 있다는 점을 감안한다면, 비교적 구하기 쉬운, 이미 출간된 논문만을 대상으로 하여 분석할 경우, 출간되지 못한 더 많은 부정적 발견들이 무시될 수 있다는 점이 문제가 된다.

### 참 고 문 헌

1. Draper NR, Smith H. Applied regression analysis. 3rd ed. New York (NY): Wiley; 1998.
2. Lang TA, Secic M. How to report statistics in medicine - Annotated guidelines for authors, editors, and reviewers. Philadelphia (PA): American College of physicians; 1997.
3. Hosmer DW, Lemeshow S. Applied logistic regression. New York (NY): John Wiley and Sons, Inc.; 1989.
4. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York (NY): John Wiley and Sons, Inc.; 1981.
5. 송혜향. 의학, 간호학, 사회과학 연구의 메타분석법. 서울: 청문각; 1998.