동의신경정신과 학회지
J. of Oriental Neuropsychiatry
Vol. 18. No. 3, 2007

동의신경정신과학회지에 발표된 임상연구논문들의 통계방법 기술 및 적용 현황

조성훈*, 황의완*, 이태림**
경희대학교 한의과대학 한방신경정신과 교실*, 한국방송통신대학교 정보통계학과**

Present Status of Description and Application of Statistics in Clinical study papers in the Journal of Oriental Neuropsychiatry.

Seung-Hun Cho*,Wei-wan Hwang*, Tae-Rim Lee**
Dept. of Neuropsychiatry, KyungHee University Korean Medical Center.*
Dept. of Information Statistics, Korean National Open University.**

Abstract

Object : This study was done to review problems in terms of statistical description and statistical application and analysis.

Methods: The authors reviewed 42 statistical clinical study papers excluding 12 Overview papers, 75 Descriptive papers, 48 Animal studies out of 177 papers in the Journal of Oriental Neuropsychiatry in the 5 years from 2002 to 2006.

Results : 1) 3 papers(7.1%) had no description of statistical method, only P-values, 25 papaers(59.5%) had tables without description of statistical method, 1 paper (2.3%) had no description of statistical method in study method.

2) 10 papers(23.8%) contained problems in terms of statistical application and analysis. 6papers (6/23, 26.0%) for Student t-test, 2 papers(2/7 28.6%)for χ^2 - test, 1 paper(1/15 6.7%) for the analysis of variance, 1 paper(1/6 16.7%) for Pearson correlation contained statistical problems.

Conclusion : It was suggested that consultation of investigators with statisticians and more extensive statistical refereeing, the form of the guidelines for description and application of statistics are needed.

key words: statistics, clinical study, description, application

---

투고일 : 10/13 수정일 : 11/4 채택일 : 11/8
교신저자 : 이태림, 서울 종로구 동숭동 169번지 정보통계학과
Tel : 02-3668-4691 Fax : 02-3668-4204 E-mail: trlee@knou.ac.kr

Ⅰ. 서 론

생명과학에 있어 통계적 자료분석은 연구결론을 얻기 위한 중요한 역할을 하게 된다. 그러나, 잘못된 통계 방법의 적용이나 분석은 연구결과 해석에 오류를 도출하여 실험의 대상자들에게나, 실험 연구자에게 막대한 노력에 대한 아무런 보상되지도 않을 뿐더러 잘못된 연구결과가 치료에 적용되었을 때, 오히려 환자에게 위험을 줄 수 있다. 또한 다른 연구가들이 잘못된 통계방법을 수정 없이 적용하여 오류가 반복될 수 있는 점도 올바른 통계방법 적용의 중요성을 나타내준다.

이에, 이번 연구는 2002년 1월부터 2006년 12월 사이에 동의신경정신과학회지에 발표된 임상연구 논문들을 대상으로 각 연구에서 각 자료를 분석하기 위하여 어떤 통계방법들이 사용되었는가, 사용된 통계방법들을 어떻게 기술하였는가를 살펴보아 그 문제점을 파악하고, 이에 대한 개선책을 제시하려는 목적으로 실행하였다.

Ⅱ. 연구 대상

연구대상은 2002년 1월부터 2006년 12월 사이에 동의신경정신과학회지에 발표된 임상연구 논문들을 대상으로 하였다.

총 177편의 논문 중, 종설 12편, 기술적인 논문(증례분석 및, 증례보고 포함) 75편, 동물실험논문 48편은 제외하고, 단순한 기술적 분석이 아닌, 통계적 분석을 사용한 임상연구 논문 42편을 본 연구의 대상으로 하였다(Table 1).

Table 1. Distribution of papers

year / categories	No. of papers					total	%
	2002	2003	2004	2005	2006		
1. Overview papers	2	4	6	0	0	12	6.8
2. Descriptive papers	7	12	16	29	11	75	42.4
3. Animal study	8	11	12	8	9	48	27.1
4. statistical clinical study	7	9	8	9	9	42	23.7
Total	24	36	42	46	29	177	100
	177						

Ⅲ. 연구 방법

Douglas G. Altman[1], 이태림 [2], 송기준[3], 조수철[4] 등이 주장한 의학연구 통계 기준을 참고로 하여 공통적이며 중요한 주장들을 중심으로 통계학적 오류 기준으로 정하여, 대상들을 살펴보았다. 평가를 하기 위한 기준은 다음과 같다.

1) 통계분석의 기술

(a)연구방법에 저자들이 통계분석에 대한 적절한 기술이 되어있는가,

(b)표나 그래프에서 대상군의 숫자, 적용된 통계방법, 결과를 명시하였는가.

2) 자료의 통계분석

(1) student의 t검정 적용의 오류

(a)대상군의 크기가 6이하에서 적용된 경우

(b)두 대상군의 크기가 차이가 2배 이상이면서 분산의 차이가 큰 경우

(c)짝짓기가 적절하지 못한 경우

(d)3군 이상에서 반복적으로 시행된 경우

(2) χ^2-검정법 적용의 오류

(a)전체 대상군의 크기가 15이하인 경우

(b)어느 한 예상기대치가 5미만에서 적용된 경우

(3) 분산분석 적용의 오류

(a)분산분석 시행 후, 사후검증이 적절하게 시행되지 않는 경우

(b)독립성, 종속성이 적절히 고려되지 않은 경우

(4) 상관성 분석 적용의 오류

(a)관측척도의 내용에 맞게 적절하게 선택되지 못한 경우

(b)상관계수를 구한 후, 이에 대한 통계적 검정이 이루어지지 않은 경우

(5) 비모수적 방법 적용의 오류

(a) 관측척도가 명칭척도 또는 순위척도가 아님

(b) 관측의 독립성 또는 종속성 또는 대상군의 수가 적절하지 않음.

Ⅳ. 연구 결과

1. 통계 방법의 기술상의 문제점

1) 임상연구논문 42편 중, 논문 전체에서 사용된 통계방법에 대한 기술이 전혀 없이 단지 P값만 기술되어 있는 논문은 3편(7.1%)이었으며,

2) 연구방법에선 특정 통계방법이 적용되었다고 기술이 되었으나 표에는 명시되어 있지 않은 논문은 25편(59.5%)이었으며,

3) 연구방법이나 표에는 사용도니 통계방법이 기술되어 있지 않고 결과란 또는 고찰란에서만 기술되어 있는 논문은 없었으며,

4) 연구방법란에는 기술이 되어 있지 않으나, 표에는 명시되어 있는 논문은 1편(2.3%)이었다(Table 2).

Table 2. status of description of statistics

Categories	Number of papers	%
1. Only P-value, no description at all	3	7.1
2. Description found in method but not found in table	25	59.5
3. Description not found in method or table but found in result or discussion	0	0
4. Description not found in method but found in table	1	2.3
Total	29	69.0

2. Student t-검정법의 적용상의 문제점들

대상이 된 42편의 논문들 중에서 23편(54.8%)에서 t-검정이 적용되었는데, 이중 6편에서 앞에서 정한 기준에 따른 문제점들이 발견되었으며, 그 내용은 다음과 같다(Table 3).

(1) 대상군의 크기가 31이상 또는 6이하에서 적용된 논문은 3편(13.0%)으로 1편은 대상군의 크기가 31이상에서 t-검정을 적용하였으며, 2편은 6이하에서 적용되었다.

(2) 두 군의 대상군의 크기의 차이가 2배 이상이면서 분산의 차이가 큰 논문은 1편이었다. 두 군의 대상군의 크기의 차이가 2배 이상인 논문이 1편이 있었으나, 분산의 차이가 크지 않아 문제점으로 지적되지 않았다.

(3) 짝짓기가 적절하지 못한 논문은 없었다.

(4) 3군 이상에서 반복적으로 t-검정을 시행한 경우는 2편(8.6%)이었다.

3. χ^2 - 검정법의 적용상의 문제점들

42편중 7편(16.7%)에서 χ^2 - 검정법이 적용

되었는데 이중 2편이 기준에 위배되었으며, 그 내용은 다음과 같다(Table 3).

(1) 교차표에서 어느 한 cell 의 이론적으로 계산된 기대치가 5미만에서 적용된 논문은 2편(4.7%)이었다.

(2) 대상군이 15이하에서 적용된 경우는 없었다.

4. 분산분석

42편중 15편(35.7%)에서 분산분석이 시행되었는데 이중 1편이 기준에 위배되었으며, 그내용은 다음과 같다(Table 3).

(1) 분산분석은 시행한 후 중다비교가 시행되지 않은 논문은 없었으며,

(2) 중다비교로서 t-검정법을 적용한 논문은 없었다.

(3) 분산분석 없이 바로 중다분석이 시행된 논문은 1편(6.7%)있었다.

(4) 관측의 독립성, 종속성이 교려되지 않은 논문은 없었다.

5. 상관성 분석

42편 중 6편(14.2%)에서 상관분석이 적용되었는데, 이중 1편(16.7%)에서 앞에서 정한 기준에 위배되는 문제점이 발견되었다(Table 3).

(1) 관측척도에 적절하게 선택되지 못한 논문은 없었으며,

(2) 상관성에서 통계적 검정을 시행되지 않은 논문이 1편 있었다.

6. 비모수적 방법

42편중 5편(11.9%)에서 비모수적 방법이 적용되었데, 앞에서 정한 기준에 위배되는 논문은 발견되지 않았다(Table 3).

(1) 관측척도가 명칭척도 또는 순위척도가 아닌 논문은 없었다.

(2) 관측의 독립성 또는 종속성 또는 대상군의 수가 적절하게 고려되지 않은 논문은 없었다.

Table 3. Numbers of papers with problems, according to problem category

Problem category	Number in subcategory	Total number(%)
1. student's t-test		
a)inadequate sample size	3	
b)equality of variance	1	6/23(26.0)
c)independence of Observation	0	
d)multiple t-test	2	
2. Chi-squared test		
a)small expected value(<5)	2	2/7(28.6)
b)small sample size(≤15)	0	
3. analysis of variance		
a)no multiple comparison	0	
b)T-test as post-hoc comparison	0	1/15(6.7)
c)No preceding ANOVA	1	
d)Independence of observation	0	
4. correlation		
a)Application in frequency	0	1/6(16.7)
b)No significance test	1	
5. nonparametric method		
a)Application in continuous data	0	0/5(0)
b)Independence of observation	0	
Total	10	10/42(23.8)

Ⅴ. 고 찰

이상의 결과에서 본 바와 같이, 2002년 1월부터 2006년 12월 사이에 동의신경정신과학회지에 발표된 임상연구 논문들에는 통계방법의 기술이나, 통계방법의 적용에 있어 문제점을 가

지고 있었다. 임상 논문 42편 중, 통계방법의 기술 면에서 문제점이 발견된 논문은 29편이었고, 통계방법의 적용 면에서 문제점이 발견된 논문은 10편에 이르고 있었다.

통계방법의 기술적인 측면에서는, 임상연구 논문 42편 중, 논문 전체에서 사용된 통계방법에 대한 기술이 전혀 없이 단지 P값만 기술되어 있는 논문은 3편(7.1%)이었고, 연구방법에선 특정 통계방법이 적용되었다고 기술이 되었으나 표에는 명시되어 있지 않은 논문은 25편(59.5%)이었으며, 연구방법란에는 기술이 되어 있지 않으나, 표에는 명시되어 있는 논문은 1편(2.3%)으로 총 29편 논문들에서 문제점이 발견되어 통계방법의 기술 측면에서는 문제가 심각하다고 볼 수 있다.

이를 볼 때, 동의신경정신과학회지에 게재되는 논문들에는 통계방법 기술에 있어 통일된 체재가 필요함을 알 수 있다. 연구자는 통계방법에 대한 정확한 이해와, 연구 계획시 적용할 통계방법을 미리 모색하고, 연구방법에 기술하는 것이 필요하다고 생각된다. 또한, 통계방법의 기술에 대하여 통일된 체재를 위해 학회지의 투고 규정에 어느 정도 규정할 필요가 있을 것이라고 생각된다.

통계방법의 적용 면에서는 Student t-검정법은 두 군간의 평균값을 비교하는 모수적 방법으로 30명 정도의 독립표본의 표본수에 적용된다. 30명을 초과하는 경우에는 Z 검정법을 적용하게 되어있다. 그리고, 작은 표본수에는 분포를 알기 어렵기 때문에 비모수 검점을 사용할 것으로 권장되고 있다. 그래서 대상군의 크기가 6이하에서 적용된 경우에는 비모수 검정을 적용하는 것이 바람직하다[4]. 또한 t-검정법의 기본 조건이 대상군의 모집단이 정규분포를 가지며, 두 모집단의 분상은 동일하다는 전제이기 때문에, 두 대상군의 크기가 차이가 2배 이상이면서 분산의 차이가 큰 경우에는 적용할 수 없다. 두군간의 관측이 종속적일 경우는 짝짓기 검정을 적용하는데, 한 표본의 서로 다른 상활에서 측정하는 경우에 적용하는 것이다. 3군 이상평균 비교에서는 분산 분석을 적용하여야 하는데, t-검정법을 반복적으로 시행하게 되면, 1종 오류가 증가하게 된다.

대상이 된 42편의 논문들 중에서 23편(54.8%)에서 t-검정이 적용되었는데, 이중 6편에서 앞에서 정한 기준에 따른 문제점들이 발견되었다. 대상군의 크기가 31이상 또는 6이하에서 적용된 논문은 3편(13.0%)이었다. 1편은 대상군의 크기가 31이상에서 t-검정을 적용하였는데, Z-검정법을 적용해야할 것으로 생각된다. 2편은 6이하에서 적용되었는데. 6이하에서는 분포를 알수 없기 때문에 비모수적인 방법을 적용해야 할 것으로 생각된다. 두 군의 대상군의 크기의 차이가 2배 이상이면서 분산의 차이가 큰 논문은 1편이었다. 이는 연구자가 분산이 차이를 나는 데이터를 다룰 때 흔히 실수할 수 있는 것으로 생각되는데, t-검정법을 사용할 때는 항상 분산의 차이를 검토해야 한다. 3군 이상에서 반복적으로 t-검정을 시행한 경우는 2편(8.6%)이었는데, t-검정법이 가장 잘 알려진 통계방법이라 생긴 오류라 생긴 오류라 생각된다. 그러나, 신경정신과 영역에서 다변량 분석을 하는 경우가 많기 때문에 연구자는 분산분석에 대한 이해를 높여야 한다고 생각된다.

χ^2-검정법은 빈도와 비율을 비교하는 통계방법으로 독립성검정과, 적합도 검정에 사용되는 통계방법이다. 이 통계방법은 많이 사용되지만, 적용시에 오류가 많은 통계중의 하나로 몇가지 조건이 필요한데, 전체 대상군의 크기가 15이하인 경우에는 오차가 커지기 때문에 그 이상이어야 하며, 한 cell의 이론적으로 계산된 기대치 중 어느 한 예상기대치가 5미만이어서는 안된다. 42편중 7편(16.7%)에서 χ^2-검정법

이 적용되었는데 이중 2편이 기준에 위배되었는데, 교차표에서 어느 한 cell의 이론적으로 계산된 기대치가 5미만에서 적용된 논문은 2편(4.7%)이었다. χ^2-검정법을 적용할 때는 연구자는 cell의 기대치를 항상 확인을 해야할 것으로 생각된다.

분산분석은 3군 이상에서 평균치를 비교하는 통계기법으로 분산분석 시행 후, 사후검증이 적절하게 시행되어야 하며, 독립성, 종속성이 적절히 고려되어야 한다. 42편중 15편(35.7%)에서 분산분석이 대부분 적절히 적용되었는데, 이중 1편이 분산분석 없이 바로 중다분석이 시행하여 중대한 문제점이 발견되었다. 이는 분산분석의 적용에 대한 이해가 시급한 것으로 생각된다.

상관성 분석은 한 변인의 다른 변인에 대한 설명력이 어느 정도인가 밝혀주는 통계기법으로 관측척도의 내용에 맞게 적절하게 선택되어야 하며, 상관계수를 구한 후, 이에 대한 통계적 검정이 이루어져야 한다. 42편 중 6편(14.2%)에서 상관분석이 적용되었는데, 이중 1편(16.7%)에서 상관성에서 통계적 검정을 시행하지 않았다. 이는 어떤 종류의 상관계수를 구하던지 그 자체는 통계적 검정력을 지니고 있지 않기 때문에 상관계수에 대한 유의성 검정이 필요한 것이다.

비모수적 방법은 표본자료가 정규분포를 따르지 않을 경우, 표본수가 적어서 분포를 알 수 없는 경우, 관측척도가 명칭척도 또는 순위척도일 경우데 적용된다. 42편중 5편(11.9%)에서 비모수적 방법이 적용되었데, 앞에서 정한 기준에 위배되는 논문은 발견되지 않았다. 이는 비모수적 방법 적용을 하는 연구자는 비교적 통계적 이해가 있는 것으로 생각된다.

이러한 결과를 통해, 통계방법 적용에 있어 문제점을 해결하는 방법을 몇 가지 모색하면 다음과 같다[4].

첫째, 연구자는 통계학자와 교류를 통해 통계학에 대한 이해를 높이며, 적용한 통계에 대하여 자문을 구한다.

둘째, 투고된 논문에 대하여 철저한 검토과정이 필요하다. 또한 투고 요령에 통계방법의 기술과 적용에 대한 지침을 마련하여, 연구자와 편집위원이 투고된 논문을 철저히 검토할 수 있도록 한다.

셋째, 논문이 발표된 후에라도 통계방법에 대한 오류가 발견되면, 동일한 실수가 반복되지 않도록 하는 제도나 과정이 필요하다.

이렇게 보다 정확한 통계방법의 기술과 적용을 추구하게 된다면, 동의신경정신과학회지와 동의신경정신과학 연구에 발전이 있으리라 생각된다.

Ⅵ. 결 론

2002년 1월부터 2006년 12월 사이에 한방신경정신과학회지에 발표된 임상연구 논문들을 대상으로 통계방법의 기술 및 적용 현황을 살펴본 결과 다음과 같은 결과를 얻었다.

1) 통계 방법의 기술적인 면에서는 통계방법의 기술 없이 단지 P 값만 기술되어 있는 논문은 3편(7.1%)이었으며, 통계방법이 표에는 명시되어 있지 않은 논문은 25편(59.5%)이었으며, 연구방법란에는 기술이 되어 있지 않으나, 표에는 명시되어 있는 논문은 1편(2.3%)이었다.

2) 42편중 10편(23.8%)의 논문에서 문제점이 발견되었다. Student t-검정법은 6편(6/23, 26.0%), χ^2 - 검정법은 2편(2/7 28.6%), 분산분석은 1편(1/15 6.7%), 상관성분석은 1편(1/6 16.7%)에서 문제점이 발견되었다.

참고문헌

1. Douglas G. Altman. Statistical reviewing for medical journals. Statistics in medicine, 1998:17:2661-2674
2. 이태림, 이재원, 박미라. 생명과학자료분석. 한국방송통신대학교출판부. 2005:37-244
3. 송기준, 한무영, 정미영, 임길섭, 김동기. 임상연구 설계 및 통계학적 분석방법. 대한비뇨기과학회지. 2005:46(8):835-841
4. 조수철, 안윤옥. 신경정신의학에 발표된 논문들의 통계방법의 기술 및 적용 현황. 신경정신의학. 1991:30(3):462-476