

의학통계학적 방법의 사용과 오류

연세대학교 의과대학 의학통계학과학교실

김동기 · 한무영 · 한혜리

Use and Misuse of Biostatistical Analysis

Dong Kee Kim, PhD, Mooyoung Han, MS and Hae Ree Han, BS

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

Objectives : This study summarizes statistical methods which have been used in the 153 original articles of the Journal of Korean Neuropsychiatric Association published in 2002 and 2003.

Methods : It deals with the appropriate statistical methods and some common errors for researchers.

Results : Among the original articles, 41 statistical uses contain errors. Many cases of errors are found in the χ^2 -test and the t-test. This study detects uses of statistical errors and suggests right statistical methods.

Conclusion : In order to improve the validity of original articles published in the Journal of Korean Neuropsychiatric Association, a more clearly stated statistical uses and closer editorial attention to statistical methods are needed. (J Korean Neuropsychiatr Assoc 2004;43 (2):141-147)

KEY WORDS : Biostatistical methods · Biostatistical errors.

서 론

최근 임상의학연구는 수집된 환자의 정보량이 방대해지고 분석내용이 복잡해짐에 따라, 이를 정확히 분석하고 해석할 수 있는 다양한 통계학적 분석방법이 요구되고 있다. 국내외 대부분의 의학학술지는 연구자가 주장하는 바를 가능한 한 통계학적 유의성을 통해서 결론을 도출하길 바라고 있으며, 이 과정에서 사용된 통계학적 방법이 타당한지가 연구 논문의 게재 여부를 결정짓는 중요한 요소가 되고 있다. 이에 따라 많은 연구자들이 그들의 연구를 수행하고 도출된 결론을 일반화시키기 위해 적절한 통계학적인 방법을 적용할 필요성이 점차 높아지고 있다. 학술지에 게재된 논문에서 통계학적 방법의 사용빈도도 점차 높아지고 있는데, 이는 비단 외국학술지만이 아니라 국내에서도 비슷한 추세를 보이고 있다. 일례로 신경정신의학회지

에 게재되는 원저에 사용된 통계학적 방법의 사용 빈도를 보면, 2002년부터 2003년까지 총 2권 12호에 게재된 원저 가운데 기술통계량 조차도 사용되지 않은 논문이 단 한건일 정도로 그 빈도는 높다고 할 수 있다. 통계학적 방법의 사용빈도가 높아지고 적절한 방법의 필요성이 증대함에 따라, 일부에서는 통계학자와의 사전 자문 혹은 통계학자들과의 공동연구를 통해 통계학적 방법의 사용범위를 확대하고 있다. 그러나 아직까지는 인력 및 여건의 부족으로 인하여 통계학적 방법이 제대로 적용되지 않거나, 분석에 대한 정확한 해석이 이루어지지 않는 등, 연구목적에 타당한 연구결과를 도출해내지 못하는 경우가 많다. 컴퓨터와 통계패키지의 발달로 연구자가 사용할 수 있는 통계학적 방법은 많아지고 그 접근성이 용이해 진 것은 사실이나, 이러한 편리함이, 적용되는 통계학적 방법에 대한 타당도를 높여주었다고 할 수는 없다.

그러므로 학술지에 게재되는 연구 논문에 대해 그 통계학적 방법의 타당성을 검증하는 일은 향후 연구자가 좋은 연구성과를 얻기 위해 정확한 통계학적 방법을 적용하는데 필요한 일이라 할 수 있다. 해외 학술지에서는 이러한 필요성에 의해 해당 학술지에 일정 기간동안 사용되었던 통계학적 방법 적용 사례를 집계하고 그 방법에 대해 설

접수일자 : 2004년 2월 1일 / 심사완료 : 2004년 3월 1일

Address for correspondence

Dong Kee Kim, Ph.D., Department of Biostatistics, Yonsei University College of Medicine, Seodaemun-gu, Shinchon-dong 134, Seoul 120-752, Korea

Tel : +82.2-361-5741, Fax : +82.2-364-8037

E-mail : dkkimbios@yumc.yonsei.ac.kr

이 논문은 보건복지부 중점공동연구지원 연구개발사업(02-PJ1-PG1-CH02-0003)의 연구비 지원을 받아 이루어졌음.

명하거나,¹⁻³⁾ 해당 학술지에 사용되었던 통계학적 방법에 대한 오류를 제시하거나,⁴⁾ 또는 의학학술지에 사용되는 통계학적 방법에 대한 지침을 제시하는 등의 방법⁵⁾ 등을 통해 연구 목적에 맞는 통계학적 방법의 올바른 사용을 위한 방안을 제시하고 있다.

본 연구는 신경정신의학회지에 게재하기를 원하는 연구자를 위하여, 본 학회지에 게재된 논문들의 통계학적 방법의 사용분야를 정리하고 흔히 발생하는 통계학적 방법의 오류사항을 정리함과 동시에 올바른 통계학적 방법의 적용을 위해 그 기준을 요약하여 제시하고자 한다.

방 법

연구대상

2002년 1월 호부터 2003년 11월 호까지 총 12권의 신경정신의학회지에 게재된 총 153건의 원저 중 통계학적 방법이 제시되고 통계량이 사용된 152건에 대해서 조사하였다.

연구방법

의학분야 연구 논문에서 주로 사용되는 통계방법을 정리하여 그 사용방법과 기준을 제시하였다. 또한 신경정신의학회지에 게재된 논문에서 사용된 통계학적 방법을 정리하여 그 사용도에 대해 조사하였으며, 이를 토대로 연구자가 흔히 범하기 쉬운 통계학적 방법의 사용 오류, 해석 오류, 또는 통계량에 대한 표기의 오류 등을 정리하였다.

결 과

통계학적 방법의 분류

Table 1은 일반적으로 의학논문에서 주로 사용되는 통계학적 방법이다. 이 방법들을 큰 범주로 묶어서 요약 정리하면 다음과 같다.

기술통계량(Descriptive Statistics)

연구 대상에 대해 그 특성을 파악하고 정의하는데 사용되고 이를 통해 그 자료의 성격을 알 수 있기 때문에 통계학적인 분석 이전에 사용된다. 주로 도수분포표와 도표 등이 사용되지만, 평균, 중앙값, 최빈값, 표준편차, 범위 등의 통계량이 적절히 활용된다.

t-검정(t-test)

두 집단의 평균의 차이를 검정하기 위해 사용된다. 일반

적으로 Student's t-test라고 하면 독립된 두 집단의 t-test (independent two-sample t-test)의 의미로 많이 사용된다. 이 경우 두 집단의 분산이 서로 같은지 다른지에 따라서 p-value가 틀려지므로 사전에 분산의 동등성 여부를 검정하여야만 한다. paired t-test는 동일한 대상에 대해 전후의 차이를 비교하는 방법이다.

분산분석(ANOVA)

셋 이상의 집단의 평균에 차이가 있는가를 검정하고자 할 경우에 사용된다. 분산분석에서 측정값에 영향을 미친다고 생각되는 효과를 요인 혹은 인자(factor)라고 하는데, 요인의 수가 하나인 경우를 일요인 분산분석(one-way ANOVA), 둘인 경우를 이요인 분산분석(two-way ANOVA)이라고 한다. 세 집단의 평균을 비교하기 위해서 t-test를 세 번 실시하는 방법은 분산분석의 가설과는 다른 검정이므로 사용하지 않도록 유의하여야 한다. 분산분석의 결과가 통계학적으로 유의하였다면, 비로소 집단간의 비교가 실시되는데, 이 절차를 다중비교(multiple comparison) 혹은 사후검정(post-hoc analysis)이라고 한다. 다중비교의 방법으로는 주로 Bonferroni 방법, Tukey 방법, 혹은 Scheffé 방법 등이 쓰인다. 공분산분석(ANCOVA)은 기저변수(예를 들면, 연령과 같은 환자의 기초변수 등)를 통제하는 분산분석법이며, 다변량분산분석(MANOVA)은 반응변수가 여러 개일 때 종속변수의 효과를 동시에 검정할 때 쓰인다. 반복측정자료의 분산분석(repeated measures ANOVA)은 어떤 실험을 함에 있어 동일한 실험대상에 대하여 실험 조건(experimental condition)을 달리하거나 또는 여러 다른 시점에서 반복적으로 특정한 값을 측정하는 경우에 시점간의 차이 혹은 집단간의 차이를 비교할 때 사용한다.

비모수 통계분석(Nonparametrics)

관찰된 자료의 수가 적은 경우, 자료가 정규분포를 가정하기 어려운 경우, 그리고 측정된 자료의 수준이 순위형인 경우에 효율적으로 사용할 수 있다. 이 경우 대표값으로는 평균보다는 중위수와 범위가 흔히 쓰인다. 윌콕슨 순위합 검정(Wilcoxon rank sum test)은 두 개의 표본이 동일한 중위수를 갖는지를 검정하기 위한 방법으로 모수적 통계분석방법 중 Student's t-test에 대응되는 비모수적 방법이다. 일반적으로 SAS(Statistical Analysis System)에서 사용하는 윌콕슨 순위합 검정과 SPSS(Statistical Package for Social Science)에서 사용하는 맨-휘트니 U 검정(Mann-Whitney U test)은 같은 결과를 제

공한다. 윌콕슨 부호순위 검정(Wilcoxon signed-rank sum test)은 모수적 통계분석 방법인 paired t-test에 대응되는 분석방법이다. 크루스칼-왈리스 검정(Kruskal-Wallis test)은 세 개 이상의 집단의 중위수가 동일한지 검정하기 위한 방법으로 모수적 통계분석 중 일요인 분산분석에 대응되는 비모수적 검정방법이다.

범주형 자료분석(Categorical Data Analysis)

분석하고자 하는 자료가 연속형 자료가 아닌 이산형 자료인 경우에 사용한다. χ^2 검정(χ^2 -test)이 대표적인 검정방법인데 $r \times c$ 형태의 분할표에서 독립성이나 동질성을 알아보고자 할 때 사용된다. 이 경우 주의할 점은 χ^2 -test는 표본수가 클 때만 적용이 가능하며, cell에 0이 포함되

Table 1. Statistical methods used in the medical journal

구분	종류	용도
기술통계량 (Descriptive statistics)	Table	Frequency table
	Graph	Histogram, bar chart
	통계량	Mean, median, mode, standard deviation, variance
t-검정 (t-test)	Student's t-test	독립적인 두 집단의 평균비교
	Paired t-test	한 집단의 전후 차이 비교
분산 분석 (ANOVA)	One-way ANOVA	요인이 하나일 때
	Two-way ANOVA	요인이 두 개일 때
	Multiple comparison 또는 post-hoc test	ANOVA의 결과가 유의한 경우 집단간을 개별 비교할 때
	ANCOVA	가져변수를 공변량으로 통제할 때
	MANOVA	종속변수가 두 개 이상일 때
비모수적 방법 (Nonparametric analysis)	Repeated measures ANOVA	반복측정자료의 변화비교
	Wilcoxon rank sum test, Mann-Whitney U test	Two-sample t-test의 비모수적 방법
	Wilcoxon signed-rank test	Paired t-test의 비모수적 방법
	Kruskal-Wallis test	One-way ANOVA의 비모수적 방법
	Friedman test	Two-way ANOVA의 비모수적 방법
	Spearman correlation	Pearson correlation의 비모수적 방법
범주형 자료분석 (Categorical data analysis)	χ^2 -test	$r \times c$ 형태의 표에서 독립성이나 동질성을 알아보기 위한 방법
	Fisher's exact test	Cell에 0이 포함되어있거나, 한 cell의 수가 5 미만일 경우에 χ^2 -test 대신 사용
	Trend test	순위형 자료의 추세를 보기 위해 사용하는 방법
	McNemar test	이분형 자료를 반복해서 관찰하였을 때, 전후의 차이가 있는지 검정할 때 사용
	Cochran-Mental-Haenszel	이분형 두 변수의 연관성을 살펴볼 때 제 3의 변수를 통제하는 방법
	Kappa index	두 집단 혹은 측정 방법의 일치도를 보기 위해 사용
상관분석 (Correlation analysis)	Pearson correlation	연속형 두 변수간의 상관성을 알아보기 위해 사용
회귀 분석 (Regression)	Linear regression	반응변수가 연속형인 자료의 회귀분석
	Logistic regression	반응변수가 이분형인 자료의 회귀분석
생존 분석 (Survival analysis)	Kaplan-Meier	비모수적 방법을 이용한 생존함수의 추정 log-rank test 혹은 Wilcoxon test로 생존함수 비교
	Cox proportional hazard model	생존시간에 대한 회귀분석방법
다변량 분석 (Multivariate data analysis)	주성분 분석 (Principal component analysis)	서로 연관이 있는 변수들이 가지고 있는 정보들을 최대한 활용하여 적은 수의 새로운 변수들을 생성하는 방법
	군집분석 (Cluster analysis)	유사한 속성을 지닌 환자군을 묶어 일정한 수의 군집으로 나누고, 각 군집의 차이점을 비교하는 방법
	판별분석 (Discriminant analysis)	독립변수가 연속형이고 종속변수는 범주형일 때 군을 분류하는 방법
	정준상관분석 (Canonical correlation analysis)	어느 특정한 환자군과 다른 관련성이 밀접한 환자군 사이의 선형적 연관관계를 파악하기 위한 분석기법
	요인분석 (Factor analysis)	관측된 변수에 근거하여 직접관측 불가능한 요인을 파악하기 위한 분석기법
표본 수 산정 (Sample size)	구조방정식 모형 (Structural equation modeling)	변수들간에 다중인과관계 및 상관관계의 추정이 필요할 때 사용
	경로분석 (Path analysis)	변수간의 인과관계를 모형도를 통해 규명하는 방법
	검정력 (Power)	집단간 차이를 밝히기 위한 최소 표본수를 산정하는 방법

어 있거나, 한 cell의 수가 5 미만일 경우에는 χ^2 -test 값 보다는 Fisher's exact test의 p-value를 적어주는 것이 올바른 방법이다. 또한 $r \times c$ 형태의 자료가 순위형(ordinal)의 성격을 가지고 있을 때는 기존의 χ^2 -test를 그대로 적용시키기보다는 순위형 변수의 증가 혹은 감소 추세(trend; 특히 선형적인 추세)에 대해 검정하는 방법이 적용될 수 있다. McNemar test는 동일 환자에 대해 반복적으로 이분형 변수를 관찰하였을 때 분석하는 기법이다. 이 때는 이분형 변수의 전후 비교에서만 사용된다. Cochran-Mantel-Haenszel test는 두 개의 이분형 변수의 연관성을 검정할 때(예를 들면, 흡연유무와 폐암유무의 연관성), 이 두 변수에 영향을 줄 수 있는 제 3의 변수(예를 들면, 음주여부)를 통제할 필요가 있을 때 적용된다. Kappa 계수(Kappa index)는 두 가지 진단 방법 혹은 측정방법의 일치도를 판단하기 위하여 사용된다. 그 일치도의 정도를 0부터 1사이의 값으로 판단하는데, 이 값이 0.75보다 크면 매우 일치도가 높은 것으로 평가하고 0.4이하이면 일치도가 낮은 것으로, 0.4에서 0.75사이인 경우는 보통정도의 일치도라고 평가한다.⁶⁾

상관분석(Correlation Analysis)

두 개의 연속형 변수들간에 어떠한 연관성이 존재하는지 알아보기 위한 방법이다. 상관관계의 척도로서는 피어슨(Pearson)이 제시한 상관계수가 대표적이다. Pearson 상관계수는 표본의 수가 크거나, 보고자 하는 변수가 정규분포를 따른다는 가정이 있어야 한다. 만약 표본의 수가 적거나, 측정 자료가 정규분포를 따른다고 가정하기 어려운 경우에는 상관분석의 비모수적인 방법인 스피어만 상관분석(Spearman correlation analysis)을 사용한다.

회귀분석(Regression Analysis)

분석하고자 하는 주변수(outcome, 혹은 종속변수)가 연속형 변수일 때, 이 주변수에 영향을 주는 변수(독립변수)와의 인과관계를 수립하고자 할 때 사용된다. 이 경우, 독립변수가 두 개 이상인 다중회귀모형(multiple regression model)이 주로 쓰이는데, 주의할 점은 독립변수들 사이에 상관관계가 높아서는 안 된다는 것이다. 또한 독립변수가 너무 많은 경우에는 유의한 변수를 선택하는 방법이 적용되는데, 전향적선택법(forward selection method), 후향적소거법(backward elimination method), 단계별 선택법(stepwise selection method) 등이 쓰인다. 이 중 단계별 선택법이 주로 쓰이는데, 이 경우 각 단계별 R^2 의 증가분

(ΔR^2)과 p-value를 표현 해주고, 최종 모형에 대한 R^2 (혹은 adjusted R^2)를 표현하여야 한다.

로지스틱 회귀분석(Logistic Regression Analysis)

회귀분석을 적용해야 하는 상황에 종속변수가 두 가지 값만을 취하는 질적 변수(예를 들면, 질병의 유/무 등)여서, 일반적인 선형 회귀분석을 그대로 적용할 수 없을 때 사용되는 방법이다. 이 때, 독립변수가 종속변수에 미치는 영향은 주로 교차비(odds ratio)로써 표현되며, 교차비가 적용될 때는 95% 신뢰구간이 같이 사용된다.

생존분석(Survival Analysis)

생존시간(survival time)에 관한 추정을 하거나 두 개 혹은 그 이상의 집단의 생존시간을 비교할 때 사용된다. 카플란-마이어(Kaplan-Meier) 추정법에 의한 생존곡선이 생존률을 추정하는데 주로 사용되고, 두 집단 이상의 생존률을 비교하는데는 log-rank 검정법과 Wilcoxon 검정법이 사용된다. 또한 생존시간에 영향을 미치는 위험인자(risk factor)를 찾기 위해서는 일반적으로 Cox의 비례위험모형(proportional hazard model)을 이용한다. 각 독립변수에 따른 생존에 미치는 위험은 상대위험도(relative risk)와 95% 신뢰구간으로 표현된다.

다변량 통계분석(Multivariate Statistical Analysis)

다변량분석(multivariate analysis)이란 연구자가 연구 대상으로부터 두 개 이상의 변수들을 측정하였을 때, 변수간의 관계를 동시에 분석할 수 있는 통계학적인 기법을 일컫는 말이다. 주성분 분석(principal component analysis)이란 한 환자에서 서로 연관이 있는 여러 변수들이 관측되었을 때, 이 변수들이 가지고 있는 정보들을 최대한 활용하여 적은 수의 새로운 변수들을 생성하는 방법이다. 군집 분석(cluster analysis)이란 한 환자에서 서로 다른 속성(attribute)을 지닌 변수를 관측하였을 때, 서로 유사한 속성을 갖는 환자들을 묶어 전체 환자를 일정한 수의 군집으로 나누는 분석 방법이다. 이렇게 나뉘어진 군집은 유사한 성격을 갖는 특성을 보이기 때문에 각 군집에 대한 내적 특징을 파악할 수 있다. 판별분석(discriminant analysis)은 분석하고자 하는 환자군이 두 개 혹은 그 이상의 집단으로 나누어져 있을 때, 이 집단을 가장 잘 분류할 수 있는 방법을 독립변수를 통하여 표현하는 방법이다. 이 때, Fisher의 판별함수(Fisher's discriminant function)가 주로 사용되며, 최종 판별결과는 정분류율과 오분류율로 표

현된다. 정준상관분석(canonical correlation analysis)이란 여러 개의 관련성이 밀접한 환자군과, 또 다른 관련성이 밀접한 환자군사이의 선형적 상관관계를 파악하기 위한 분석기법이다. 요인분석(factor analysis)이란 한 환자에서 차원이 서로 상의한 변수를 관측하였을 때, 여러 개의 변수들이 서로 어떻게 연결되어 있는가를 분석하여 이들 변수간의 관계를 공동요인(내재적 차원)을 이용하여 더욱 밀접한 상관관계를 보이는 변수들끼리 묶어 줌으로써 그 관계를 보다 간결하고 명확하게 해 주는 분석 방법이다. 이 때 요인을 줄이는 기준으로 아이겐 값(eigenvalue)이 사용되며 축약되어 생성된 변수의 설명력이 얼마인지는 각 요인의 설명된 분산의 크기로 파악하게 된다.

구조방정식 모형 분석(Structural Equation Modeling)

이 방법은 다차원의 구조로 얻어진 변수들간에 다중 인과관계 및 상관관계의 추정이 필요할 때 사용된다. 구조방정식을 통하여 직접 변수로써 관측되지 못한 단일 개념 혹은 복합개념에 대하여 추정할 수 있으며, 이 추정과정에서 발생할 수 있는 오차에 대한 설명이 가능하다. 경로분석(path analysis)은 변수간의 인과관계를 모형도(path diagram)를 통해 규명하는 방법이다. 이 때 주의할 점은 가능한 한 유의한 변수만을 선택하여 모형을 구축해야 하며 최종 모형의 적합도에 관한 검정통계량 등이 제시되어야 한다.

Table 2. Uses and errors of statistical methods in the Journal of Korean Neuropsychiatric Association

분석방법	사용건수	오류
Two-sample t-test	49	1
Paired t-test	16	1
χ^2 -test	59	25
Fisher's exact test	6	0
One-way ANOVA	39	2
Nonparametric one-way ANOVA	18	2
Factorial ANOVA	8	1
Repeated measures ANOVA	10	1
MANOVA	5	1
ANCOVA	12	1
Correlation analysis	44	2
Linear regression	20	1
Logistic regression	11	1
Survival analysis	1	0
Principal component analysis	5	0
Factor analysis	11	1
Cluster analysis	2	0
Discriminant analysis	7	0
Path analysis	3	1
계	326	41

신경정신의약학회에 사용된 통계적 방법의 적용 유형

본 연구에서 검토된 원저에 사용된 통계학적 방법은 총 326건이며 그 빈도는 Table 2와 같다. 사용된 방법을 살펴보면 두 집단을 비교하는 two-sample t-test와 paired t-test가 많이 사용되었으며 비율을 비교하는 χ^2 -test와 Fisher's exact test가 많이 사용되었다. 두 집단 이상을 비교하는 분산분석법으로는 ANOVA, factorial ANOVA, ANCOVA, MANOVA와 nonparametric ANOVA 등이 활용되었다. 회귀와 상관분석법으로는 Pearson correlation, linear regression, 그리고 logistic regression 등이 고루 사용되었다. 다변량 분석법으로는 discriminant analysis와 cluster analysis 등이 활용되었고, 그밖에 survival analysis 등이 활용되었다.

오류 건수를 살펴보면 χ^2 -test에서 가장 많은 오류를 보였으며 그 밖의 통계학적 방법에서도 한 두건의 오류를 보여, 326건 중에 41건이 사용된 방법 중에 오류를 포함하고 있었다.

신경정신의약학회에 사용된 통계적 방법의 오류 사례

Table 3은 2002년도와 2003년도에 투고된 논문의 통계오류 사례를 정리한 표이다. 가장 많은 빈도를 차지하고 있는 통계오류 사례는 χ^2 -test의 경우인데, cell에 포함된 수가 적어서 Fisher's exact test를 적용해야 함에도 불구하고 χ^2 -test를 그대로 적용한 경우이다. t-test의 경우에는 Student's t-test와 paired t-test를 혼동한 경우가 있었다. 예를 들면, 실험 전후의 차이를 비교하는 자료에 paired t-test 대신에 Student's t-test를 사용한다거나, 그 반대의 경우로 대조군과 환자군의 비교에 paired t-test를 사용하는 경우 등이다. 또한 빈도수의 비교에 χ^2 -test 대신에 t-test를 사용하는 오류도 있었다.

ANOVA에서는 repeated measures ANOVA를 사용해야 하는 자료에 independent t-test를 반복해서 사용한다거나 통계분석 방법의 ANOVA와 사후검정(post-hoc analysis)의 표기를 혼동하는 오류, 두 집단에 비교에서는 불필요한 사후검정 방법을 사용했거나 'Two-way ANOVA'라고 표기하지 않고 '2×2 변량 분석'이라고 표현하는 용어사용의 부적절성이 발견되었다. 상관분석은 자료의 수가 적은 경우임에도 Spearman correlation을 적용하지 않고 모수적 방법인 Pearson correlation을 사용하는 사례가 있었으며, r값에 대해서 부호에 따라 해석해야 됨에도 불구하고 모두 '정적 상관'이라고 해석하는 오류도 있었다. 비모수 자료의 분석 부분에서는 뚜렷한 근거를 제시하지 않은 상태에서의 비모수적 방법을 적용하거나 자

Table 3. Detail descriptions of statistical errors in the Journal of Korean Neuropsychiatric Association

구분	유형
t-test	1) t-test를 사용함에 있어서 자료의 특성에 따라 independent한 경우와 paired한 경우의 구분이 모호함. - 실험 전후의 차이를 비교하는 자료에 paired t-test 대신에 independent t-test 사용 - 대조군과 환자군의 비교에 paired t-test 사용 2) 빈도수의 비교에 χ^2 -test 대신에 t-test 사용 3) 전체에 대한 평균값과 그 차이를 비교한 값의 표기 누락 4) 용어 사용 부적절 : 'two sample t-test'대신에 'two tailed'로 잘못 기술
χ^2 -test	cell에 0이 포함되어있거나, 한 cell의 수가 5 미만일 경우임에도 Fisher's Exact test의 p-value를 사용하지 않고 χ^2 -test 값을 적용
ANOVA	1) repeated measures ANOVA를 사용해야 하는 자료에 independent t-test를 반복해서 사용 2) 통계분석 방법의 ANOVA와 사후검정 (post-hoc analysis)의 표기 혼동 3) 두 집단에 비교에서는 불필요한 사후검정 방법 사용 4) 용어 사용 부적절 : 'two-way ANOVA'라고 표기하지 않고 '2×2 변량 분석'이라고 표현
Canonical correlation	1) 자료의 수가 적은 경우임에도 Spearman correlation을 적용하지 않고 모수적 방법인 Pearson correlation을 사용 2) r값의 부호에 따라 해석해야 됨에도 불구하고 모두 '정적 상관'이라고 해석
Nonparametric method	1) 뚜렷한 단계를 제시하지 않은 상태에서의 비모수적 방법적용 2) 자료가 세 개 이상의 그룹인 경우임에도 Kruskal-Wallis 검정이 아닌 두 집단을 비교하는 Mann-Whitney U 분석이나 Wilcoxon rank sum test를 반복 적용
Path analysis	경로분석을 적용할 경우, 통계량과 경로도만 명시하는 경우가 있는데, 경로분석은 그 결과에 대해서 모형의 적합도를 판정해야 함에도 모형에 대한 χ^2 -test에 대한 통계량과 p-value 혹은 Goodness Fit Index (GFI)에 대한 표기 및 해석 누락
기 타	1) 분석 결과를 table로 명시하는 경우, 보통 t값, d.f, F값 등을 명시하고 p-value값은 명시하지 않음 2) 본문에서 결과 해석시 생략해도 무방한 t값, d.f, F값 등의 표기만 되어있고, 소수점 세 자리까지 써 주어야 하는 p-value 값을 누락하는 경우가 있었으며, p<0.0001로 표기해야 하는 p-value를 통계패키지에서 나온 표기법 그대로 p=0.0000이라고 표기하는 경우가 많았다. 3) 0.05 미만으로 또는 일정하게 명시되어야 할 유의수준이 분석에 따라 변화하여 자의적으로 해석 4) 통계패키지에서 나온 표기법 그대로 p=0.0000이라 표기 (이는 p<0.0001이라고 표현해야 함) 5) 통계분석방법에서 기술한 내용과 세부결과에서 기술한 내용에 용어 선택을 혼동하거나 그 내용이 상이함

료가 세 개 이상의 그룹인 경우임에도 불구하고 Kruskal-Wallis 검정이 아닌 두 집단을 비교하는 Mann-Whitney U 분석이나 Wilcoxon rank sum test를 반복 적용하는 사례가 있었다. 경로분석을 적용할 경우, 통계량과 경로도만 명시하는 경우가 있는데, 경로분석은 그 결과에 대해서 모형의 적합도를 판정해야 함에도 모형에 대한 χ^2 -test에 대한 통계량과 p-value 혹은 적합도지수(Goodness of Fit Index : GFI)에 대한 표기 및 해석을 누락하였다.

이 외에도 통계량을 표로 작성하는 부분에서 오류가 많이 발생하였다. 예를 들면, 분석 결과를 표로 명시하는 경우, 생략해도 무방한 t값, 자유도(degree of freedom : d.f), F값 등의 표기만 되어있고, 소수점 세 자리까지 써 주어야 하는 p-value 값을 누락하는 경우가 있었으며, p<0.0001로 표기해야 하는 p-value를 통계패키지에서 나온 표기법 그대로 p=0.0000이라고 표기하는 경우가 많았다.

통계학적 방법 부분에서는 0.05 미만으로 또는 일정하게 명시되어야 할 유의수준이 분석에 따라 변화하여 자의적으로 해석하는 경우가 있었으며, 통계분석방법에서 기술한 내용과 세부결과에서 기술한 내용에 용어 선택을 혼동하거나 그 내용이 상이한 경우도 있었다.

이러한 오류는 통계학적 분석방법의 사용 용도와 해당 분석방법의 적용에 대한 정확한 인식 부족에서 기인한 것

으로 분석방법에 대한 좀 더 세밀한 이해가 필요하다 할 수 있겠다.

결론

본 연구에서는 의학연구에서 주로 사용되는 통계분석 방법의 종류와 올바른 사용방법에 대해 정리하였다. 또한 신경정신의학회지에 2002년도와 2003년도에 게재된 논문에서 사용된 통계방법을 검토하여 주로 사용된 통계적 분석방법의 사용현황과 그 오류에 대해 살펴보았다. 신경정신의학회지에 게재된 거의 대부분의 원저가 통계학적 분석방법을 사용하고 있으며 그 결과를 통해 연구자가 주장하고자 하는 결론에 도달하고 있었다. 하지만 좋은 자료나, 자료에서 얻어낸 통계학적 결과를 갖고도 326건의 통계학적 방법의 적용 건수 중에 41건이 통계학적 방법의 정확하지 못한 적용이나 미비한 해석으로 연구의 성과를 완벽하게 얻어내지 못하는 점은 아쉬운 부분이라 하지 않을 수 없다. 하지만, 통계학을 전공하지 않은 연구자의 입장에서 모든 통계학적 방법과 그에 따른 특성을 정확히 파악하는 것이 결코 쉬운 일이 아님은 주지의 사실이다. 그러나 본 연구에서 제시한 통계학적 방법의 적용기준을 통해 연구자가 자료의 특성을 정확히 파악하여 수행하고자 하는 연

구에 알맞는 통계적 분석 방법을 사용할 수 있고, 또한 그에 대해서 올바른 결론을 내릴 수 있다면 연구결과에 대해서 더 정확한 해석을 할 수 있으며 그에 따른 올바른 결론을 내릴 수 있을 것이다. 나날이 발전하는 통계분석 패키지 등을 통해 편리하게 통계량을 얻을 수는 있지만, 통계학적 분석방법을 정확히 적용하지 못하거나, 통계학적 분석방법을 적용은 하였으나 결과를 정확히 해석하지 못한다면 좋은 의학연구에 도달할 수 없을 것이다. 본 연구에서 통계학적 분석을 정리하고 사용오류를 제시하는 것은 연구자가 자료 분석과 해석을 하는데 있어 조금이나마 도움이 되기 위한 것이다. 상위 등급의 학술지일수록 통계학적 분석방법의 사용과 그 해석 및 결과의 표현방법에 대해 높은 관심을 보이기 때문에 본 연구의 이러한 기준 제시는 향후 신경정신의학회지의 좋은 연구성과를 위해서도 반드시 필요한 작업이라고 판단한다.

중심 단어 : 의학통계학 방법 · 의학통계학 오류.

REFERENCES

- 1) Emerson JD, Colditz GA. Statistical in Practice. N Engl J Med 1983; 309:709-713.
- 2) Chinn S. Statistics for European Respiratory Journal. Eur Respir J 2001;18:393-401.
- 3) MacArthur RD, Jackson GG. An Evaluation of the Use of Statistical Methodology in the Journal of Infectious Disease. J Infect Dis 1984; 149:349-354.
- 4) White SJ. Statistical Errors in Papers in the British Journal of Psychiatry. Br J Psychiatry 1979;135:336-342.
- 5) Bailar JC III, Mosteller F. Guidelines for Statistical Reporting in Articles for Medical Journals: Amplifications and Explanations. Ann Intern Med 1988;108:266-273.
- 6) Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977a;33:159-174.
- 7) Hair JF, Anderson RE, Tatham RL. Multivariate data analysis, 2nd ed., NY: Macmillan Publishing Company;1987.
- 8) Rosner B. Fundamentals of Biostatistics, 5th ed., Duxbury;2000.