

## 대한한방부인과학회지에서 사용된 통계방법에 관한 연구

<sup>1</sup>가천대학교 한의과대학 한방부인과학교실, <sup>2</sup>가천대학교 경상대학 응용통계학교실  
김윤상<sup>1</sup>, 오현숙<sup>2</sup>, 임은미<sup>1</sup>

### ABSTRACT

#### A Review of Statistical Methods in the Journal of Oriental Obstetrics & Gynecology

Yoon-Sang Kim<sup>1</sup>, Hyun-Sook Oh<sup>2</sup>, Eun-Mee Lim<sup>1</sup>

<sup>1</sup>Dept. of Gynecology, College of Oriental Medicine, Ga-Chon University

<sup>2</sup>Dept. of Applied Statistics, College of Business & Economics, Ga-Chon University

**Objectives:** The purpose of this article is not until to investigate the changes and types of statistical methods and to point out the statistical errors after analyzing the method of articles that improve the quality of the statistical analysis of papers published in the Journal of Oriental Obstetrics and Gynecology.

**Methods:** Papers published in the Journal of Oriental Obstetrics and Gynecology from 2009 to 2011 were reviewed for methodological and statistical validity using a modified version of Ahn's checklist. A statistician reviewed individual papers and evaluated the list items in the checklist for each paper. To avoid the potential assessment error by the statistician who lacks expertise in the field of Oriental Obstetrics and Gynecology.

**Results:** A total of 190 papers including 64 original articles, 40 reviews article, 58 case report and 28 brief communication were reviewed. Statistics methods used in 121 papers were composed of t-test(58.7%), ANOVA test(19.8%) and  $\chi^2$ -test (14.0%) et al. Whereas only 14.9% of papers were free of statistical errors, the number of omission errors was 58 and the number of commission errors was 149 each.

**Conclusions:** A variety of statistical errors were encountered in papers published in the Journal of Oriental Obstetrics and Gynecology. Accordingly, researchers should be more careful when it comes to describing and applying statistical methods.

**Key Words:** Oriental Obstetrics and Gynecology, Statistical Error, Statistics, Statistical Methods, Oriental Medicine

"이 논문은 2011년도 가천대학교 교내 연구비 지원에 의한 결과임(KWU-2011-R363)."

## I. 서 론

오늘날 다양한 학문 분야에서 자료의 수집, 수집된 자료의 정리와 분석 및 학술연구를 위해 통계기법을 발전시켜왔고, 한의학을 비롯한 의학, 치과학, 약학 및 간호학 등의 의학계열에서도 질병의 역학, 치료의 효용성, 약물의 안정성 및 부작용 등에 관한 수많은 연구들을 수행하고 있다<sup>1,2)</sup>.

연구자들은 자신들의 실험, 임상 및 기타 연구 결과를 정확하게 분석하기 위하여 통계기법을 활용하는 것은 매우 중요함에도 불구하고 우리나라에서 뿐만 아니라 역학이 발달한 영국을 비롯한 선진국의 많은 의학 관련 학술지에 게재된 논문에서도 통계기법 사용의 오류가 적지 않다고 하였다<sup>3-6)</sup>.

미국의학협회지에서는 의학연구에서의 통계학적 방법론이 차지하는 중요성을 인식하여 게재신청 논문의 통계전문가에 의한 논문 심사 과정을 의무화하였고, 게재된 논문들에 대한 통계적 오류의 사후평거나 개선 방안들이 잇따르고 있는 실정이다<sup>6,7)</sup>.

국내 의학계에서도 1990년대 이후 대한마취과학회를 비롯한 다양한 전문학회를 중심으로 해당 학회지의 통계적 실태를 분석하는 연구가 이루어졌고<sup>8-12)</sup>, 한의학에서도 대한침구학회 등을 중심으로 유사한 노력들이 있어왔지만 대한한방부인과학회지의 통계기법 사용에 대한 평가는 아직 이루어지지 않았다<sup>13)</sup>.

이에 저자는 대한한방부인과학회지에 게재된 학술논문의 질적 제고와 신뢰의 향상은 매우 중요하기 때문에 본 연구에서는 대한한방부인과학회지 게재 논문의 통계 오류 현황을 파악하고 이에 근거한 개선 방안을 마련하고자 본 연구를 시행하게 되었다.

## II. 연구대상 및 방법

### 1. 연구대상

본 연구를 위해 2009-2011년 사이에 출판된 대한한방부인과학회지에 수록된 논문들을 연구대상으로 하였는데, 매년 각 1권 4호씩 발간되어 3년간 총 3권 12호이고, 등재된 논문의 수는 총 190편이었다.

### 2. 연구방법

각각의 논문에 적용된 통계기법의 적정성 평가는 통계점검표를 활용하였는데 1997년 국제의학학술지 편집위원회에서 제시한 통계적 서술원칙 15개항에 비해 비교적 간략한 안 등<sup>14)</sup>이 제시한 것을 바탕으로 한 박 등<sup>3)</sup>의 통계점검표를 위주로 수정, 보완하여 사용하였다(Table 1).

통계점검표는 통계 전문가와 종합 토의 및 문헌 고찰을 통하여 결정하였고, 통계점검표에 포함된 평가 항목은 연구의 종류, 적용된 통계기법의 종류 및 통계기법 적용의 타당성 등이다.

Table 1. Checklist for Evaluation of Statistical Validity for Papers in the Journal of the Oriental Obstetrics and Gynecology

연도/권(호)/페이지 :	
논문제목 :	
저자 :	
1. 연구종류	(1) 원저 (2) 중설 (3) 임상 및 증례보고 (4) 단신보고
2. 통계기법	(1) $\chi^2$ -test (2) Fisher's extract test (3) t-test (4) Paired t-test (5) ANOVA (6) Repeated measures ANOVA (7) Mann-Whitney U test (8) Kruskal-Wallis test (Wilcoxon rank sum test) (10) Pearson correlation analysis (9) Wilcoxon signed rank test (12) Linear regression (11) Spearman correlation analysis (14) Survival analysis (15) Etc (13) Logistic regression
3. 통계기법 적용의 타당성	(1) 생략 오류 ① 기초자료의 불충분한 기술 ② 사용된 통계기법의 불충분한 기술 ③ 검정력 및 신뢰구간의 불충분한 기술 ④ 사용된 통계기법에 대한 타당성의 불충분한 검증 (2) 시행 오류 ① 대표치 및 산포도의 부적절한 기술 ② 잘못된 분석방법의 사용 ㉠ 보정 없이 t-test의 반복 사용 ㉡ 변수의 독립성을 고려하지 않은 t-test ㉢ 불충분한 표본 수에서 $\chi^2$ -test 사용 ㉣ 분산분석 후 적절한 다중비교가 고려되지 않은 경우 ㉤ 비모수적 자료를 모수적으로 분석 ㉥ 자료의 성격과 맞지 않는 통계기법 사용 ③ 중복 시행한 가설검정 ④ 부당하게 유도된 결론 ㉦ 표본에 대한 통계처리없이 모집단에 대한 결론을 유추 ㉧ 통계결과를 확장하여 비논리적 결론을 유도한 경우 ㉨ 통계용어의 잘못된 사용 ㉩ 연구결과를 일반 집단에 적용하면서 그 한계를 언급하지 않은 경우 (3) 통계적 오류 없음

(1) 연구 논문의 종류  
대한한방부인과학회지에서 사용하고 있는 분류에 따라 원저, 중설, 임상 및 증례보고 및 단신보고 등으로 구분하였다.

(2) 통계 처리의 유무  
평가 대상 논문이 통계적인 방법을 사용하고 있는지의 여부를 조사하고, 통계기법을 사용한 논문을 대상으로 해당 통계 방법의 빈도를 측정하였는데, 만약

한 논문에서 여러 가지 통계기법이 사용된 경우 이를 각각의 횟수로 계산하였다.

(3) 통계기법 적용의 타당성 조사  
통계 오류는 같은 논문에 서로 다른 오류가 두 가지 이상 있는 경우 이를 각각 횟수로 측정하였고, 같은 오류가 2회 이상 반복된 경우 1회로 횟수를 계산하였다.

통계기법 적용의 타당성 항목에 분류

한 통계 오류는 ‘생략 오류’와 ‘시행 오류’로 크게 구분하였는데 ‘생략 오류’는 기초자료의 불충분한 기술, 사용된 통계 기법의 불충분한 기술, 검정력 및 신뢰 구간의 불충분한 기술 및 사용된 통계 기법에 대한 타당성의 불충분한 검증 등으로 구분한 반면에 ‘시행 오류’는 대표치 및 산포도의 부적절한 기술, 잘못된 분석방법의 사용, 중복 시행한 가설검정 등 부당하게 유도된 결론 등으로 구분하였다.

(4) 통계점검표 작성 및 점검

통계 전문가가 개별 논문을 대상으로 통계점검표를 작성하였는데, 각 논문은 대한한방부인과학회 홈페이지에서 다운로드하여 사용하였고, 일차 평가자가 한방부인과 전문가가 아니어서 발생할 수

있는 평가 오류를 고려하여 개별 통계점검표는 한방부인과 교수 2인이 각각 한 차례 추가적 세부 점검을 하였다.

### III. 결 과

2009년부터 2011년까지 대한한방부인과학회지에 게재된 총 190편의 평가 대상 논문은 원저 64편, 종설 40편, 증례보고 58편 및 단신 28편 등으로 구성되었는데, 이 중 121편의 논문에서 통계기법을 사용한 반면에 나머지 69편은 통계기법을 전혀 사용하지 않은 66편과 통계기법의 선택 없이 자동 프로그램에 의해 처리된 3편이었다(Table 2).

Table 2. The Type of Research and the Statistical Method Used in the Journal of Oriental Obstetrics and Gynecology

	Year			Total (n=190)
	2009 (n=72)	2010 (n=59)	2011 (n=59)	
Type of Research				
Original Article	28	21	15	64(33.7%)
Reviews Article	16	14	10	40(21.1%)
Case Report	19	16	23	58(30.5%)
Brief Communication	9	8	11	28(14.7%)
Statistics(No Statistics)	46(26)	39(20)	36(23)	121(63.7%)

통계기법을 사용한 121편의 논문에서 사용된 통계기법의 횟수는 t-test 71회 (58.7%), ANOVA test 24회(19.8%),  $\chi^2$ -test 17회(14.0%) 및 Paired t-test 12회

(10.0%) 등의 순이었고, 그 외 기타 사용된 통계기법으로는 LSD test, Tukey HSD test, Duncan test, Cluster analysis 및 Levene test 1회 등이 있었다(Table 3).

Table 3. The Incidence of Inferential Statistics Used in the Journal of the Oriental Obstetrics and Gynecology from 2009 to 2011

Statistical Method	Year			Total (n=121)
	2009 (n=46)	2010 (n=39)	2011 (n=36)	
$\chi^2$ - test	4	9	4	17(14.0%)
Fisher's exact test	1	0	0	1(1.0%)
t-test	27	21	23	71(58.7%)
Paired t-test	6	3	3	12(10.0%)
ANOVA test	12	7	5	24(19.8%)
Repeated measures ANOVA test	0	0	0	0(0.0%)
Mann-Whitney U test(Wilcoxon rank sum test)	3	5	3	11(9.1%)
Kruskal-Wallis test	1	1	3	5(4.1%)
Wilcoxon signed rank test	0	2	0	2(1.7%)
Pearson correlation analysis	4	2	3	9(7.4%)
Spearman correlation analysis	1	0	6	7(5.8%)
Linear regression	3	0	0	3(2.5%)
Logistic regression	1	2	0	3(2.5%)
Survival analysis	0	0	0	0(0.0%)
Others	3	4	4	11(9.1%)

대상 논문 중 통계 적용의 오류가 없는 논문은 18편(14.9%)인 반면에 중복 산정한 생략 오류와 시행 오류는 각각 58회, 149회로 측정되었다. 통계기법의 오류를 세부적인 항목으로 살펴보면 '자료의 성격과 맞지 않는 통계기법 사용'

이 51편(42.1%), '보정 없이 t-test의 반복 사용'이 36편(29.8%), '사용된 통계기법의 불충분한 기술'이 32편(26.4%) 및 '비모수적 자료를 모수적으로 분석'이 19편(15.7%) 등이 비교적 자주 발생하는 오류로 관찰되었다(Table 4).

Table 4. Number of Articles with Statistical Errors in the Journal of the Oriental Obstetrics and Gynecology

Type of statistical error	Year			Total (n=121)
	2009 (n=46)	2010 (n=39)	2011 (n=36)	
1. Errors of omission				
① Incomplete description of basic data	5	5	2	12(10.0%)
② Incomplete description of statistical test performed	15	10	7	32(26.4%)
③ Incomplete description of statistical power	0	0	0	0(0.0%)
④ Incomplete verification of statistical method	9	4	1	14(11.6%)
2. Errors of commissioning				
① Inadequate description of measures of central tendency or dispersion	0	0	2	2(1.7%)
② Incorrect analysis				
- repeatedly used t-test without correction	17	9	10	36(29.8%)
- t-test not considering variable independency	0	0	6	6(5.0%)
- Chi-square test with inappropriate sample size	0	0	1	1(1.0%)

- ANOVA without appropriate multiple comparison	3	1	2	6(5.0%)
- analysis with parametrical method for non-parametrical data	7	4	8	19(15.7%)
- failure to select proper statistical method	21	18	12	51(42.1%)
③ Multiplicity on hypothesis testing	0	0	0	0(0.0%)
④ Unwarranted conclusion				
- with no statistical test	5	4	2	11(9.1%)
- illogical conclusion	5	4	4	13(10.7%)
- inadequate statistical terminology	1	1	2	4(3.3%)
- extrapolation of study results	0	0	0	0(0.0%)
3. No statistical error	5	4	9	18(14.9%)

## IV. 고찰

대한한방부인과학회지의 수록된 논문의 비율은 1994년까지는 문헌논문이 대부분을 차지하였으나, 1995년부터 실험논문이 급격하게 증가하기 시작하여 문헌논문보다 많은 비중을 차지하기 시작한 반면에 임상논문의 경우 1996년부터 점차 증가하기 시작하여 2001년 이후로 가장 많은 비중을 차지하였다<sup>15)</sup>.

이러한 실험논문과 임상논문의 증가는 다양한 통계기법의 활용이 필수적으로 수반되어야 하는데 의학논문에 사용된 통계적 기법의 타당성에서는 첫째, 연구 대상이 표본일 경우 통계적 검정을 통해 일반화된 결론이 유도되었는지, 모집단일 경우에는 통계적 검정의 필요 유무에 대해 평가하고, 둘째, 결론을 유도하는데 있어서 적절한 통계검정방법이 적용되었는지를 평가하며, 마지막으로 방법 부분에 기술된 대상자의 수, 특징과 표에 제시된 값들이 일치하고, 변수의 분포에 맞는 통계량이 사용되었는지가 중요하다<sup>16)</sup>.

통계기법 사용의 오류는 학술 논문의 신뢰성을 현저하게 떨어뜨리고 잘못 분석된 연구 결론을 토대로 반복적으로 적용되는 임상 진료가 이루어지는 중대한

오류를 일으킬 수 있으므로 절대로 간과해서는 안된다<sup>3,17)</sup>.

본 연구는 통계기법 적용의 중요성을 고취시키기 위하여 2009년부터 2011년까지 3년간 발간된 총 190편의 논문을 분석하였는데, 원저 64편, 종설 40편, 증례보고 58편 및 단신 28편 등으로 구분되었고, 이 중 121편(63.7%)의 논문에서 통계기법을 사용한 반면에 나머지 69편은 통계기법을 전혀 사용하지 않은 66편과 통계기법의 선택 없이 자동 프로그램에 의해 처리된 3편이었으며, 이는 대한방사선종양학회지 63% 및 대한응급의학회지 69.1%와 비슷한 분포를 보였다<sup>3,10,12)</sup>.

통계기법을 사용한 121편의 논문에서 사용된 통계기법의 횟수는 t-test 71회(58.7%), ANOVA test 24회(19.8%),  $\chi^2$ -test 17회(14.0%) 및 Paired t-test 12회(10.0%) 등의 순이었고, 그 외 기타 사용된 통계기법으로는 LSD test, Tukey HSD test, Duncan test, Cluster analysis 및 Levene test 1회 등이 있었는데, 다른 의학 관련 학회지에 게재된 논문들과 비슷한 형태로 나타났다<sup>8,10,17)</sup>.

대상 논문에 대한 통계기법의 오류에 관한 분석에서는 통계기법의 오류가 없는 논문은 18편(14.9%)으로 이는 BMJ

의 52%, 대한이비인후과학회지의 63.4% 및 한방재활의학과과학회지의 80% 등에 비해 통계기법의 오류 비중이 높게 나타난 것이고, 대한의학협회지의 97.6%, 대한성형외과학회지의 94.5% 및 결핵 및 호흡기 학회지의 87%에 비해서는 낮게 나타난 것이다<sup>10-11,17)</sup>.

또한, 중복 산정한 생략 오류와 시행 오류는 각각 58회, 149회로 측정되었는데, 이는 대한응급의학회지 게재 논문의 생략 오류와 시행 오류가 각각 70.6%, 29.6%로 나타난 것과는 상이한 결과로 개별 연구 결론의 신뢰성에 영향을 미칠 수 있을 뿐만 아니라 대한한방부인과학회지의 질적 평가에 영향을 준다는 점에서 개선책을 마련해야 할 것으로 사료된다<sup>12)</sup>.

개별 항목에서 살펴보면 ‘자료의 성격과 맞지 않는 통계기법 사용’ 51편(42.1%)과 ‘보정 없이 t-test의 반복 사용’ 36편(29.8%)이 가장 많이 나타났는데, 대부분 세 집단 이상에서 ANOVA test를 시행해야함에도 불구하고 두 집단씩 묶어서 t-test를 반복 적용한 것으로 이런 경우 쌍별 t-test에서 설정한 유의수준은 전체 집단 간 비교 시 설정한 유의수준보다 크게 되어 유의성 검정이 정확하지 않게 되므로 주의해야 한다.

만약, ANOVA 모형을 적용할 경우 t-test의 자료가 정규성과 등분산성을 만족하면서 동시에 설명력( $R^2$ )을 기술하여 모형의 타당성을 보여줘야 하는데, 이는 회귀분석에서도 동일하고, 로지스틱 회귀모형 분석에서는 적합도 검정(Goodness of Fit)을 통한 모형의 타당성 입증에 필요하다.

또한, 질적자료에 대한  $\chi^2$ -test를 실

행할 때도 표본수가 50 이상으로 충분히 커야하는데 특히 15 이하일 때는 오차가 커지기 때문에 반드시 Fisher's exact test로 처리해야 하고, 실험논문에서도 실험군 수가 충족되지 않았을 때는 비모수적 방법의 사용과 정규성 검정을 반드시 실시해야 한다.

그 외 ‘사용된 통계기법의 불충분한 기술’ 32편(26.4%)과 ‘비모수적 자료를 모수적으로 분석’ 19편(15.7%) 등의 오류가 관찰되었는데, 표본수의 부정확한 기술과 도표나 그림의 불충분한 통계 설명이 많았고, 특히, 연구방법에서 사용되었다고 기술한 통계기법이 실제로 사용되지 않은 경우도 있었는데 이는 연구윤리에 위배되므로 더욱 심사에서 지적된다면 저자에게 주의를 주어야 할 것으로 사료된다.

또한, 통계적 검증을 거치지 않고 기술적 통계치만으로 연구에 대한 결론을 유도하거나 논문 주제에 비해 너무 단순한 통계기법만을 사용함으로써 나온 결론을 유도할 수 못하는 경우도 관찰되었다.

연구자가 통계에 관한 기본 지식 없이는 자료로부터 통계적 추론을 통해 올바른 결론에 도달하기 어렵고, 통계전문가도 연구의 논점을 이해하지 못한 채 자료의 분석만을 시행할 때는 또다른 오류를 범할 수 있으므로 상호간의 적절한 교류를 통해 해결해야 한다.

또한 본 연구에서는 간략한 통계점검표만으로 분석하였기 때문에 오류의 경중을 구분하지 못하였는데, 향후 학회지 투고 과정에서 통계 심사를 강화하고, 무엇보다 회원들을 대상으로 꾸준한 교육을 병행함으로써 향후 대한한방부인과

학회지의 질적 강화를 위해 더 많은 노력을 도모해야 할 것으로 생각된다.

- 투 고 일 : 2012년 2월 1일
- 심 사 일 : 2012년 2월 1일
- 게재확정일 : 2012년 2월 7일

## V. 결 론

2009년부터 2011년까지 대한한방부인과학회지에 게재된 총 190편의 논문을 분석하여 논문에 사용된 통계기법과 그 오류에 대해 조사한 결과 다음과 같은 결론을 얻었다.

1. 2009년부터 2011년까지 대한한방부인과학회지에 게재된 총 190편의 논문 중 121편의 논문에서 통계기법을 사용하였다.
2. 논문에서 사용된 통계기법은 t-test(58.7%), ANOVA test(19.8%),  $\chi^2$ -test(14.0%) 및 Paired t-test(10.0%) 등의 순으로 나타났다.
3. 통계적 오류가 없는 논문은 18편(14.9%)인 반면에 중복 산정한 생략 오류와 시행 오류는 각각 58회, 149회였고, 세부적인 오류는 '자료의 성격과 맞지 않는 통계기법 사용'(42.1%), '보정 없이 t-test의 반복 사용'(29.8%), '사용된 통계기법의 불충분한 기술'(26.4%) 및 '비모수적 자료를 모수적으로 분석'(15.7%) 등의 순으로 관찰되었다.

이상의 연구 결과 대한한방부인과학회지의 질 관리 측면에서 투고 저자들의 노력, 통계 오류에 관한 적절한 심사 및 주기적인 모니터링이 필요하리라 사료된다.

## 참고문헌

1. 이준영. 한눈에 알 수 있는 의학통계학. 제2판. 서울:E PUBLIC. 2007:3-5, 14-21.
2. 임희정. 대한치과교정학회지와 미국교정학회지에서 사용된 통계기법의 비교분석 및 고찰(1999-2003). 대한치과교정학회지. 2004;34(5):371-9.
3. 박희철 등. 대한방사선종양학회지 게재 논문의 통계적 오류 현황. 대한방사선종양학회지. 2008;26(4):289-94.
4. Gore SM, Jones IG, Rytter EC. Misuse of statistical methods:critical assessment of articles in BMJ from January to March 1976. Br Med J. 1977;1:85-7.
5. Avram MJ et al. Statistical methods in anesthesia articles:an evaluation of two American journals during two six-month periods. Anesth Analg. 1985; 64:607-11.
6. White SJ. Statistical errors in papers in the British journal of psychiatry. Br J Psychiatry. 1979;135:336-42.
7. Lee HJ. Medical research and the use of statistical methods. Seoul:Korea Medical Book Publisher. 2000.
8. 고흥 등. 대한마취과학회지에 게재된 논문의 통계적 분석에 관한 고찰(1981년-1990년). 대한마취과학회지. 1993; 26(1):22-7.
9. Park IK et al. An Evaluation of the



- Statistical Techniques used in the 2995-1996 editions of the Korean Journal of Clinical Pathology. Korean J Clin Pathol. 1999;19:460-4.
10. 신명희 등. 이비인후과 영역에서의 통계적 기법의 활용에 대한 고찰. 대한이비인후과학회지. 1999;42:1364-9.
  11. 황건 등. 대한성형외과학회지 게재논문의 통계적 오류현황. 대한성형외과학회지. 2001;28:302-9.
  12. 유태영 등. 대한응급의학회지에 게재된 논문의 통계기법에 대한 고찰. 대한응급의학회지. 2003;14:475-80.
  13. 이승덕. 대한침구학회지 논문의 통계적 오류에 관한 연구. 대한침구학회지. 2004;21(1):176-88.
  14. Ahn YO, Le HK. Development of A Checklist for Assessing the Methodological and Statistical Validity of Medical Articles. Korean J Med Educ. 1991; 3:19-35.
  15. 윤영진 등. 대한한방부인과학회지 수재 논문 분석. 대한한방부인과학회지. 2011;24(2):97-103.
  16. Park YG. Evaluation for validity of statistical method. J Korean Acad Fam Med. 2006;27:93-6.
  17. 박태용, 허태영, 신병철. 한방재활의학과회지의 통계적 오류에 관한 고찰( I ). 한방재활의학과회지. 2010; 20(4):105-30.