

대한치과보철학회지에 게재된 논문의 통계적 오류: 국문논문과 영문논문의 비교

박동규¹ · 최용근² · 김영수² · 신상원^{2*}

고려대학교 임상치의학대학원 고급치과보철학과 ¹대학원생, ²교수

연구목적: 대한치과보철학회지에 사용된 통계적 방법의 통계적 오류현황을 파악하여 통계적 방법의 잘못된 활용으로 인한 치의학 연구논문에서의 문제를 검토하고, 또한 통계적 오류의 유형과 빈도를 파악하여 동일문제의 재발 방지 및 효율적 대처의 기회로 삼고자 함이다.

연구 재료 및 방법: 2000년부터 2006년까지의 대한치과보철학회지에 게재된 총 399편의 논문 중 통계적 방법이 사용된 292편의 논문을 조사했고, 이를 위해 1997년 국제의학학술지 편집위원회("생의학 학술지에 투고하는 원고의 통일양식의 통계학적 방법론에 관한 서술지침" 15개항을 기초로 하여 1) 통계 프로그램의 불명시, 2) 통계 방법의 부적절한 기술, 3) 통계용어의 오용 등으로 크게 3가지 범주로 구성된 점검표를 사용하여 통계적 방법의 종류 및 통계적 오류부분을 확인하고 이를 한글판과 영문판으로 비교해 보았다. 그 결과는 다음과 같다.

결과 및 결론: 1. 통계 프로그램 명시 오류는 한글판에선 54%가, 영문판에선 59%가 있었으나 통계적으로 유의한 차이를 보이지 않았다 ($P = .66$). 2. 통계 방법의 부적절한 기술에 관한 오류는 한글판에선 49%가, 영문판에선 33%가 있었으나 통계적으로 유의한 차이를 보이지 않았다 ($P = .09$). 3. 통계용어의 오용 오류는 한글판에선 54%가, 영문판에선 24%가 있었으며 통계적으로도 유의한 차이가 있었다 ($P = .01$). 상기 결과로 볼 때 통계용어의 오용 오류 이외엔 통계적 유의성을 보이지 않아서 한글판과 영문판의 통계적 오류의 비율은 유사하게 나타난 결과로 추정되며, 전체적으로 대한치과보철학회지의 통계적 타당성은 다소 부족하게 나타났다. 따라서 향후 치의학 연구자는 통계적 방법을 기술하고 적용할 때 통계적 오류를 피하기 위해 주의를 기울여야 할 것이며, 아울러 치의학 논문에서의 통계학적 방법론에 대한 인식의 제고가 필요하다고 보인다. (*대한치과보철학회지 2009;47:273-85*)

주요단어: 통계적 방법, 통계적 오류, 치의학 연구, 대한치과보철학회지

서론

치의학 연구에서 연구의 목적을 효과적으로 달성하기 위해서는 연구 설계를 타당하게 구상하여야 할 뿐 아니라, 연구에서 얻어진 자료로부터 합리적이고 객관적인 결론을 끌어낼 수 있도록 적절한 통계적 방법을 적용하여야 한다. 통계적 방법의 잘못된 활용은 연구결과에 심각한 문제를 일으킬 수도 있고 연구논문의 타당성은 통계적 방법의 타당성에 의해 이루어지므로, 치의학 연구 시에는 통계적 방법을 효율적으로 사용하여 통계적 오류를 피하도록 해야 할 것이다.^{1,2}

이러한 적절한 통계적 방법의 적용과 통계적 오류에 관한 문제는 외국의 유명 의학학술지에서도 예외가 아닌 실정이어서, 게재된 논문의 70 - 80% 이상에서 통계적 방법을 사용하고 있으며, 그 중 40 - 70%에서 통계적 오류가 발견된다고 한다.³ 또한 국내 의학학술지를 대상으로 한 여러 연구에서도 게재된 논문의 30 - 90%가 통계처리 방법을 사용하고 있으며, 그 중 통계적 오류가 적게는 60 - 80%, 많게는 95 - 100%에 이르는 것으로 보고되고 있다.^{4,5}

또한 치의학 연구결과를 발표하고 보고하기 위해서 뿐만 아니라, 발표된 연구결과를 치의학 학술지를 통해 접하는 독자 입장에서 연구자가 사용한 연구방법 및 통계적 방법에 대한 지식이 없이는 연구자의 연구결과와 여기서 얻어진 결론을 효과적으로 평가하기 어렵다. 물론 치의학 연구를 위한 통계학적 방법론에 대한 정확한 이해를 위해서는 별도의 노력과 공부 필요하겠지만 모든 부분에서 통계전문가 같은 지식을 갖추기엔 한계가 있다. 따라서 통계적인 식견이 있는 연구자라고 하더라도 치의학 연구논문 작성을 위해 통계분석을 시행할 경우에는 통계전문서적과 더불어 유사한 분석방법을 활용한 선행연구논문을 활용할 것이다.

이렇듯 의학 연구논문에서는 제시하는 의학적 결론보다 연구방법론이 더 큰 관심대상이 되기도 한다는 점에서도 의학논문에서 사용된 통계적 오류를 줄이려는 노력이 여타 국내외 의학학술지에서는 꾸준히 지속되어 오고 있다. 미국의학협회지에서는 의학연구에서의 통계학적 방법론이 차지하는 중요성을 인식하여 이미 지난 1965년부터 게재신청 논문의 통계전문가에 의한 논문심

교신저자: 신상원

136-705 서울시 성북구 안암동 5가 126-1 고려대학교 임상치의학대학원 고급치과보철학과 02-818-6874; e-mail, swshin@korea.ac.kr

원고접수일: 2009년 5월 19일 / 원고최종수정일: 2009년 6월 17일 / 원고채택일: 2009년 7월 2일

사과정을 의무화하였으며,⁴ 미국심장학회지, 미국안과학회지, 영국정신의학회 등 기타 유명 학술지에서도 해당 학술지에 게재된 논문들의 통계적 오류를 사후평가하거나 그 개선 방안을 제시하는 연구들이 1970년대 이후로 잇따르고 있다.⁶ 이러한 추세에 따라 1988년 국제의학 학술지 편집위원회에서는 생의학 학술지에 투고하는 원고의 통일양식 (Uniform Requirements for Manuscripts Submitted to Biomedical Journals) 개정 3판에서 통계학적 방법론에 관한 서술지침을 포함하기에 이르렀고, 1997년에는 "생의학 학술지에 투고하는 원고의 통일 양식" 개정 5판에서는 통계적 서술원칙 15개항을 제시하였다.⁷ 국내에서도 1970년대 이후로 대한예방의학회,⁸ 대한마취과학회,⁹ 대한안과학회,¹⁰ 대한성형외과학회¹¹ 등에서 해당 학술지에 게재된 논문들의 통계적 오류를 평가하는 연구들이 이루어져왔다.

이러한 의학학술지 게재논문들의 통계적 오류 평가가 예전엔 주로 의학통계학자들에 의해 시작되었으나 최근에 와서는 통계에 대한 식견이 있는 의학연구자들에 의해 이루어진 연구결과들도 많이 보고되고 있다. 이들 연구들은 일반적으로 평가대상인 통계적 오류의 범위가 넓지는 않지만 대신에 의학연구에서 활용빈도가 많은 특정한 분석기법 활용의 적절성과 의학통계의 주요 문제를 다루는 경우가 많으며, 이는 기존에 흔히 관찰되는 통계적 오류의 유형과 빈도를 파악하는데서 시작하여 이를 동일문제의 재발 방지 및 효율적인 대처의 기회로 삼고 있다.⁴ 그럼에도 불구하고 현재 국내 치의학 관련 학술지에선 이러한 통계적 오류의 유형과 빈도를 파악하여 치의학 연구논문에서 사용된 통계적 오류를 줄이려는 노력이 여타 국내외 의학학술지에 비해 미미한 실정이다.

이에 본 연구에서는 2000년 영문판 학회지의 발간 이후 대한치과보철학회지의 한글판과 영문판을 비교하여 통계적 방법이 사용된 논문의 빈도와 사용된 통계적 방법의 종류 및 통계적 오류의 현황을 살펴보고 그 타당성을 검토하여 보았다. 아울러 전체 대한치과보철학회지의 한글 논문과 영문 논문에서의 해당 통계적 오류의 비율을 추정해 본 결과 다소의 지견을 얻었으므로 이에 보고하는 바이다.

연구 대상 및 방법

1. 연구 대상

이미 작성된 논문에 대한 통계적 오류의 파악은 그 나

름의 의미는 있겠으나 발표한 후 시간이 많이 경과한 후에는 통계적 오류에 관심을 두고 해당 논문을 읽는 연구자가 많지도 않을 테고 논문에 사용된 자료의 원형을 찾기가 쉽지 않을 수도 있어 실제적으로는 큰 도움이 되지 않을 것이므로,⁶ 본 연구에서는 연구 대상의 범위를 최근 2000년 영문판 학회지의 발간 이후 사용된 통계적 방법의 종류와 그 활용에 통계적 오류가 없는지 영문판과 한글판을 비교하여 살펴보고자 대한치과보철학회의 도움을 받아 2000년부터 2006년까지의 대한치과보철학회지의 총 399편의 논문 중 통계적 방법을 사용한 292편의 논문을 조사하였다.

2. 연구 방법 및 점검표

1997년 국제의학학술지 편집위원회의 "생의학 학술지에 투고하는 원고의 통일양식의 통계학적 방법론에 관한 서술지침" (Uniform requirements for manuscripts submitted to biomedical journals: Statistics) 15개항을 기초로 하고, 통계적 오류와 관련된 논문을 다수 저술하고 자주 인용된 Gore,¹⁷ 안,¹⁸ 최,⁵ Altman,³ 박,¹⁹ 이⁴ 등의 점검표 및 저서를 참고하고 정리하여, 통계적 오류와 관련된 논문에서 공통적으로 자주 언급되었고 통계전문가가 아닌 사람도 발견할 수 있는 통계적 오류 위주로 1) 통계 프로그램의 불명시, 2) 통계 방법의 부적절한 기술, 3) 통계 용어의 오용 등으로 크게 3가지 범주로 구성된 점검표를 사용하여 통계적 오류부분을 확인하였다. 특히 두 번째의 통계적 오류 범주인 통계 방법의 부적절한 기술에선 1991년에 안과 이¹⁸가 개발한 "의학연구논문의 방법론 및 통계처리기법의 타당성 평가를 위한 점검표" 중 일목요연하게 정리된 통계처리기법의 타당성 평가 항목을 활용하였다. 아울러 각 세부항목의 평가지침은 저자가 제시한 기준을 따랐다. 이 점검표는 이 등²⁰이 재조사법에 의한 관찰자내 신뢰도에서 kappa치가 모든 항목에서 0.7 - 0.9로 상당한 일치를 보여 주었다고 하였다.

물론 평가의 성격상 평가자 개인의 관점 및 평가능력에 따라 영향을 받을 수밖에 없는 한계가 있는 것도 사실이나, 본 연구의 목적은 게재 논문에서 사용된 통계처리 기법의 종류와 활용 빈도 및 이의 활용이 적절한 지 그 현황을 파악하여 통계적 오류를 줄이기 위한 노력의 필요성을 언급하고자 함이므로, 이를 위해 통계적 오류와 관련된 논문에서 공통적으로 자주 언급되었고 통계전문가가 아닌 사람도 발견할 수 있는 통계적 오류 위주로 (1) 대표값 및 산포도에 대한 부적절한 기술, (2) 부적합한 분석방법, (3) 중복 시행한 검정 등으로 크게 3가지 범주로

하여 반복 점검하였다. 본 연구에서 대한치과보철학회지를 연도별로 점검한 점검표의 항목을 정리하면 다음과 같다.

1. 통계 사용 논문의 빈도
(Frequency of the articles using statistics)
2. 통계 처리기법의 빈도
(Frequency of the statistical techniques)
3. 통계적 오류의 빈도
(Frequency of the statistical errors)
 - 1) 통계 프로그램의 불명시
(Unspecified computer statistical packages)
 - 2) 통계 방법의 부적절한 기술
(Inadequate description of statistical methods)
 - (1) 대표값 및 산포도에 대한 부적절한 기술
(Inadequate description of measure of central tendency or dispersion)
 - (2) 부적합한 분석방법
(Incorrect analysis)
 - (3) 중복 시행한 검정
(Multiplicity on hypothesis testing)
 - 3) 통계 용어의 오용
(Misuse of statistical terms)

3. 통계 처리

Microsoft Excel 2003 (Microsoft, Seattle, WA, USA)을 이용하여 한글판과 영문판을 비교하여 연도별로 자료를 정리하고 기술하였다. 그리고 통계용 컴퓨터 프로그램인 dBSTAT for Windows version 4.5 (dBSTAT, Seoul, Korea)를 사용하여 전체 대한치과보철학회지 한글판과 영문판에서의 통계적 오류 비율을 95% 신뢰구간으로 모수추정하였으며, 전체 대한치과보철학회지의 한글판과 영문판에서 통계적 오류의 비율의 차이를 95% 신뢰구간으로 추정하고 Chi-square test로 Yates' correction for continuity (연속성 수정)을 하여 유의수준 0.05에서 통계 검정하였다.

결과

1. 통계 사용 논문의 빈도

2000년부터 2006년까지의 대한치과보철학회지의 총 399편의 논문 중 통계처리기법을 사용한 292편의 논문에서 한글판과 영문판 각각의 연도별 빈도는 Table I 과 같다. 한글판에서는 총 284편의 논문 중 212편 (75%)이 통계

적 방법을 사용하였으며, 영문판에서는 총 115편의 논문 중 80편 (70%)이 통계적 방법을 사용한 논문이었다.

2. 통계처리기법의 종류와 빈도

동일 논문에서 여러 가지 통계처리기법을 사용한 경우도 각각을 모두 포함시켜 살펴보니, 대한치과보철학회지에서 통계적 방법을 사용한 전체 292편의 논문 중에서 한글판 212편의 논문에 사용되어진 통계처리기법의 종류와 그 연도별 빈도는 Table II와 같았으며, 영문판 80편에 사용되어진 통계처리기법의 종류와 그 연도별 빈도는 Table III과 같았다. 가장 기본적인 것이라 할 수 있는 통계처리기법인 *t*-test 및 실험설계와 주로 관련된 ANOVA (분산분석, Analysis of Variance)와 특별한 분포나 가정에 큰 영향 없이 사용가능한 비모수 통계방법 (Non-parametric statistics)이 주로 사용되었으며, 그 외에는 상관 분석 (Correlation Analysis), 회귀분석 (Regression Analysis) 등이 있었다. 한글판에서는 통계처리기법의 명시가 불명확한 경우까지 포함한 전체 301건의 통계처리기법 중에서 ANOVA가 171건 (57%), *t*-test가 49건 (16%), 비모수 통계방법이 43건 (14%)으로 다빈도를 차지하였으며, 영문판에서는 전체 109건의 통계처리기법 중에서 한글판에서와 마찬가지로 ANOVA가 61건 (56%), *t*-test가 17건 (16%), 비모수 통계방법이 20건 (18%)으로 다빈도를 보였다. 한편, 사용되어진 비모수 통계기법으로는 한글판과 영문판 공히 Kruskal-Wallis test, Mann-Whitney U test가 그 중 다수를 차지하였으며, 그 외에 Wilcoxon test, Chi-square test 등이 있었다. 2005년도 들어서 한글판에서는 MANOVA (Multivariate analysis of variance) 같은 다변량 분석 방법이 나타나기 시작했으나 영문판에서는 전무하였다.

그리고 국제의학학술지 편집위원회의 “생의학 학술지에 투고하는 원고의 통일양식 중에 통계학적 방법론에 관한 서술지침”에서도 연구 설계나 통계학적 방법에 대한 참고문헌은 통계 이론의 원전보다는 잘 설계되고 표준적으로 실행한 연구 보고를 (페이지와 함께) 인용하는 것이 좋다고 언급하였다. 이는 통계학적 방법론을 활용한 연구논문 발표 시엔 연구방법론의 과학적 근거를 제시 한다는 입장에서 그렇고, 의학적 결론 자체보다도 연구방법론에 대한 참고 문헌이 독자나 후속연구자들에게 더 관심일 수 있다는 점에서도 참고문헌을 가능한 구체적으로 기재해야 할 것이다. 즉, 연구방법을 활용할 때 문헌 고찰한 참고문헌을 기록하는 것과 마찬가지로 통계적 기법과 원리에 대한 참고문헌을 밝히는 것도 당연하게 여겨야 할 것이다. 그럼에도 불구하고 Weibull

Table I. Frequency of the articles using statistics in the Journal of Korean Academy of Prosthodontics by year

| | | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | Total |
|-----------------|---------------------------|------|------|------|------|------|------|------|-------|
| Korean version | Total articles | 46 | 36 | 32 | 40 | 40 | 42 | 48 | 284 |
| | Articles using statistics | 37 | 27 | 26 | 31 | 27 | 34 | 30 | 212 |
| English version | Total articles | 20 | 17 | 13 | 16 | 18 | 16 | 15 | 115 |
| | Articles using statistics | 4 | 13 | 11 | 13 | 15 | 12 | 12 | 80 |

Table II. Frequency of the statistical techniques used in the Korean version of the Journal of Korean Academy of Prosthodontics by year

| Statistical technique | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | Total |
|-----------------------|------|------|------|------|------|------|------|-------|
| t-test | 11 | 5 | 4 | 5 | 8 | 10 | 6 | 49 |
| ANOVA | 34 | 15 | 22 | 30 | 17 | 21 | 32 | 171 |
| Correlation | 1 | | 1 | 3 | | 6 | 1 | 12 |
| Regression | 1 | | 1 | 3 | | 1 | 2 | 8 |
| Non-parametric | 11 | 2 | 6 | 6 | 2 | 7 | 9 | 43 |
| Multivariate | | | | | | 3 | 3 | 6 |
| Others * | 3 | | 1 | 1 | | 3 | 1 | 9 |
| Undefined | 1 | | | | 2 | | | 3 |
| Total | 62 | 22 | 35 | 48 | 29 | 51 | 54 | 301 |

*Kaplan-Meier survival analysis, interrater reliability etc.

Table III. Frequency of the statistical techniques used in the English version of the Journal of Korean Academy of Prosthodontics by year

| Statistical technique | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | Total |
|-----------------------|------|------|------|------|------|------|------|-------|
| t-test | 1 | | 3 | 1 | 3 | 5 | 4 | 17 |
| ANOVA | 4 | 8 | 8 | 12 | 12 | 9 | 8 | 61 |
| Correlation | | 1 | | | | 1 | 3 | 5 |
| Regression | | | | | 1 | | 2 | 3 |
| Non-parametric | 2 | 8 | 1 | 3 | 2 | 3 | 1 | 20 |
| Multivariate | | | | | | | | |
| Others * | | | 1 | | | 1 | 1 | 3 |
| Undefined | | | | | | | | |
| Total | 7 | 17 | 13 | 16 | 18 | 19 | 19 | 109 |

*K-means clustering analysis etc.

분석과 Kaplan-Meier 생존분석, 그리고 동의도 test라고 기술한 평가자간 신뢰도분석 (interrater reliability)에 관한 참고문헌 뿐, 전체 논문 중 단 3편이 대한치과보철학회지에 서의 통계적 기법과 원리에 대한 참고문헌의 전부였다.

3. 통계적 오류의 빈도

2000년부터 2006년까지의 대한치과보철학회지의 총 399편의 논문 중 통계처리기법을 사용한 한글판 212편과 영문판 80편의 전체 292편의 논문에서 통계적 오류항목을 크게 3가지 범주로, 즉 1) 통계 프로그램의 불명시, 2) 통계방법의 부적절한 기술, 3) 통계 용어의 오용 등으로 나누어 점검하였다. 영문판과 한글판을 비교하여 살펴보고자 전체 대한치과보철학회지 한글판과 영문판에서의 통계적 오류 비율을 95% 신뢰구간으로 모수 추정하였으며, 전체 대한치과보철학회지의 한글판과 영문판에

서 통계적 오류의 비율의 차이를 95% 신뢰구간으로 추정하고, Chi-square test로 Yates' correction for continuity 을 하여 유의수준 0.05에서 통계 검정한 결과는 다음과 같으며 Table IV로 정리하였다.

1) 통계 프로그램의 불명시의 경우가 한글판에선 54%, 영문판에선 59%로 나타났으나, 그 차이는 95% 신뢰구간으로 모수추정 시 -13%에서 22%였으며 Chi-square test 후 Yates 수정 결과도 통계적 유의성은 없었다 [$\chi^2(1, N=292) = 0.195, P > .05$].

2) 통계방법의 부적절한 기술의 경우가 한글판에선 48%, 영문판에선 33%로 나타났고, 그 차이는 95% 신뢰구간으로 모수추정 시 3%에서 28%였으나 Chi-square test 후 Yates 수정 결과는 통계적 유의성이 없었다 [$\chi^2(1, N=292) = 2.841, P > .05$].

Table IV. Frequency of the statistical errors by the comparison between Korean version and English version in the Journal of Korean Academy of Prosthodontics

| Category of statistical errors | Korean version | | English version | | χ^2 value (P-value) |
|---|-------------------------|----------------|-------------------------|----------------|-----------------------------|
| | No.(%) | Articles using | No.(%) | Articles using | |
| | 95% Confidence Interval | statistics | 95% Confidence Interval | statistics | |
| Unspecified computer statistical packages | 115 (54) 0.48 - 0.61 | 212 | 47 (59) 0.43 - 0.75 | 80 | 0.195 (P = .66) |
| Inadequate description of statistical methods | 102 (48) 0.41 - 0.55 | 212 | 26 (33) 0.22 - 0.43 | 80 | 2.841 (P = .09) |
| Misuse of statistical terms | 114 (54) 0.47 - 0.60 | 212 | 19 (24) 0.14 - 0.33 | 80 | 10.369 (P = .01)* |

Note. χ^2 (Chi-square) tests with Yates's correction for continuity were done at $\alpha = 0.05$ level. (degrees of freedom = 1, N = 292)

*P < .05, statistically significant.

3) 통계 용어의 오용의 경우가 한글판에선 54%, 영문판에선 24%로 나타났으며, 그 차이는 95% 신뢰구간으로 모수추정 시 19%에서 42%였고 Chi-square test 후 Yates 수정 결과도 통계적 유의성이 있었다 [$\chi^2(1, N=292)=10.369, P < .05$].

대한치과보철학회지 영문판이 정식 단행본으로 나온 것은 2000년도부터이지만 그 이전에도 부정기적으로 영문 논문은 발표되었었다. 따라서 모수추정은 가능하다고 본다. 하지만 Statistical Analysis System (SAS)가 1976년에, Statistical Package for the Social Sciences (SPSS)가 1968년에 개발되었으며, 1980년대 중반 이후에 와서야 통계용 컴퓨터 프로그램이 개인 컴퓨터용으로 개발되었으니 통계용 컴퓨터 프로그램 명시 오류항목은 상용화된 1990년대 이후를 모수추정의 해당 시점으로 보아야 할 것이다. 다

시 말해 1990년대 이전의 논문을 예외로 본다면 전체 대한치과보철학회지에서도 국문논문과 영문논문에서 이와 유사한 비율의 오류가 있을 것이라 추정된다. 실제로 대한치과보철학회지에서는 1996년 제34권 1호에서 처음으로 통계용 컴퓨터 프로그램이 언급되었다. 각 항목별로 살펴본 통계적 오류의 세부내용은 다음과 같다.

1) 통계 프로그램의 불명시

국제의학학술지 편집위원회의 “생의학 학술지에 투고하는 원고의 통일양식의 통계학적 방법론에 관한 서술지침”에서도 언급했듯이, 일반적으로 사용하는 컴퓨터 프로그램도 명시해야한다. 대한치과보철학회지의 연도별로 명시된 통계용 컴퓨터 프로그램은 Table V, VI에 정리하였으며, 사용된 통계용 컴퓨터 프로그램의 종류

Table V. Frequency of the computer statistical packages used in Korean version of the Journal of Korean Academy of Prosthodontics by year

| Statistical packages | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | Total |
|-----------------------|------|------|------|------|------|------|------|-------|
| SPSS | 10 | 7 | 9 | 14 | 9 | 13 | 9 | 71 |
| SAS | 4 | 1 | 1 | 3 | 4 | | 4 | 17 |
| Microsoft Excel | | 1 | 1 | | 2 | 1 | | 5 |
| Statgraphics plus | | | | | 1 | | | 1 |
| SigmaStat for Windows | | 1 | | | 1 | | | 2 |
| Minitab | | | | | | 1 | | 1 |
| Total | 14 | 10 | 11 | 17 | 17 | 15 | 13 | 97 |

Table VI. Frequency of the computer statistical packages used in English version of the Journal of Korean Academy of Prosthodontics by year

| Statistical packages | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | Total |
|-----------------------|------|------|------|------|------|------|------|-------|
| SPSS | 2 | 4 | 2 | 2 | 5 | 4 | 8 | 27 |
| SAS | | 1 | | 1 | | | 2 | 4 |
| Microsoft Excel | | | | | 1 | | | 1 |
| Statgraphics plus | 1 | | | | | | | 1 |
| SigmaStat for Windows | | | | | | | | |
| Minitab | | | | | | | | |
| Total | 3 | 5 | 2 | 3 | 6 | 4 | 10 | 33 |

는 SPSS (73%), SAS (20%), Microsoft Excel (4%), Statgraphics plus (1%), SigmaStat for Windows (1%), Minitab (1%) 순으로 많이 사용되었다. 한글판에서는 통계적 방법을 사용한 총 212편의 논문 중에서 97편 (46%)만이, 영문판에서는 통계적 방법을 사용한 총 80편의 논문 중에서 33편 (41%)만이 명시되어 있었다. 다시 말해 Table IV에서와 같이 한글판 115편 (54%)과 영문판 47편 (59%)에 선 구체적으로 명시되어 있지 않았다. 비록 2000년부터 2006년까지의 대한치과보철학회지에 사용된 통계 프로그램이 불명시된 경우가 평균적으로 영문판의 경우가 약간 더 높게 나타났으나 Table I을 참고로 하여 Table V, VI에서 보는 것처럼 2006년도에는 한글판에선 통계 프로그램이 불명시된 경우가 30편중에 17편 (57%)인 반면 영문판에선 12편중에 2편 (17%)뿐으로 최근 들어 더 개선됨을 보였다.

2) 통계 방법의 부적절한 기술

국제 의학학술지 편집위원회의 “생의학 학술지에 투고하는 원고의 통일양식의 통계학적 방법론에 관한 서술지침”에서도 언급했듯이, 연구 내용을 이해할 만한 독자가 연구 데이터를 확보했을 때 논문에 나타난 결과를 확인할 수 있을 정도로 통계학적 방법을 자세하게 기술하여야 한다. 이에 안과 이¹⁸가 개발한 통계처리기법의 타당성 평가 항목을 활용하여 (1) 대표값 및 산포도에 대한 부적절한 기술, (2) 부적합한 분석방법, (3) 중복 시행한 검정 등으로 3가지 범주로 나누어 살펴본 결과, Table IV에서처럼 총 292편의 논문 중에서 통계학적 방법의 부적절한 기술의 오류가 한글판에선 102편 (49%), 영문판에선 26편 (33%)으로 전체 128편 (44%)의 논문에서 발견되었다. 그 세부 내용은 다음과 같았으며 Table VII로 정리하였다.

(1) 대표값 및 산포도에 대한 부적절한 기술

대표값과 산포도는 변수의 척도와 표본 수, 자료의 정규성 여부 및 적용하고자하는 통계기법과의 일관성을 고려하여서 표나 그림으로 기술통계량을 나타내는 경우, 연속변수는 대표값과 산포도를 함께 제시하되 정규분포를 할 때 평균과 표준편차, 정규분포가 아닐 때는 중앙값과 범위가 사용되어야 한다. 이는 연구 결과의 타당성을 직접적으로 저해하진 않지만 연구 자료에 대한 기초적인 정보전달이 불충분하게 되는 문제가 생길 수 있기 때문이다.²¹² 그런 점에서 볼 때 대한치과보철학회지에서는 표본수가 적고 정규분포가 의심스러운데도 이에 대한 언급 없이 비모수 검정법을 사용까지 해놓고선 중앙값과 범위로 기술되지 않고 평균과 표준편차로 자료를 정리하고 언급한 경우가 있었다. 또한 표준편차 (SD, Standard Deviation)와 표준오차 (SEM, Standard Error of the Mean)를 혼동해서 사용하면 안 된다. 측정 관찰된 표본에서의 각각의 값이 평균값을 중심으로 흩어져있는 정도를 나타내는 표준편차 대신 같은 모집단에서 동일 연구를 무수히 반복한다는 가정 하에서 얻는 표본자료에서 모집단에 관한 추론 과정 중에 사용하는 추정치인 표준오차를 연구 자료의 요약목적으로 사용하는 것은 잘못이다.^{4,12,21} 그럼에도 불구하고, Mean ± SD가 아니라 Mean ± SEM으로 서술된 경우도 있었다.

이러한 대표값 및 산포도에 대한 부적절한 기술의 오류가 Table VII에서처럼 2000년부터 2006년까지의 대한치과보철학회지의 한글판에서는 통계적 기법을 사용한 총 212편의 논문 중에서 21편 (10%)에, 영문판에서는 통계적 기법을 사용한 총 80편의 논문 중에서 2편 (3%)에 있었다. Table IV에서처럼 통계방법의 부적절한 기술의 오류가 발견된 128편의 논문 중에서는 한글판과 영문판을 합쳐서 23편 (18%)이 이에 해당하였다.

Table VII. Frequency of the inadequate description of statistical methods in the Journal of Korean Academy of Prosthodontics by year

| category | version | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | Total |
|--|---------|------|------|------|------|------|------|------|-------|
| Inadequate description of central tendency or dispersion | Korean | 4 | 2 | 1 | 3 | 3 | 4 | 4 | 21 |
| | English | | 1 | | | | 1 | | 2 |
| Incorrect Analysis | Korean | 10 | 5 | 6 | 7 | 6 | 8 | 5 | 47 |
| | English | 1 | | 4 | 4 | 2 | 3 | 1 | 15 |
| Multiplicity on hypothesis testing | Korean | 8 | 3 | 7 | 4 | 5 | 3 | 4 | 34 |
| | English | 2 | 2 | | 1 | | | 4 | 9 |
| Total | Korean | 22 | 10 | 14 | 14 | 14 | 15 | 13 | 102 |
| | English | 3 | 3 | 4 | 5 | 2 | 4 | 5 | 26 |

(2) 부적합한 분석방법

t 검정과 같이 모집단의 확률 분포 및 이로부터 유도된 표본분포를 이용하는 통계적 가설검정을 모수검정 (母數檢定, parametric test)이라고 부르며, 모수검정은 모집단의 확률분포를 중요시하게 취급한다. 따라서 일반적인 모수검정은 "모집단이 정규분포를 할 때" 등과 같은 전제조건을 붙이게 되며, 이러한 전제조건인 가정이 만족되어야만 검정의 타당성을 보장할 수 있다. 즉, 사용하고 자 하는 통계분석에서 기본적으로 요구하는 조건과 가정들을 만족하는 지에 따라 분석 결과의 타당성이 정해지므로 통계처리방법의 가정을 만족시키지 못하면서 무리하게 적용하면 안 된다. 따라서 각 통계처리방법의 가정에 대해 최소한의 사항을 반드시 살펴보아야 할 것이다.^{21,22} 하지만 대한치과보철학회지 논문에서는 이러한 정규성 검정을 위한 Kolmogorov-Smirnov test는 K-S test라는 용어로 6건, 등분산성 검정을 위한 Levene's test 언급은 단 한 건 뿐이었다. 그런데도 표본수가 적거나 정규성이 의심됨에도 불구하고 이에 관한 언급 없이 동일 자료를 모수 검정인 t -test와 비모수 검정인 Mann-Whitney U test를 혼용하거나 모수 검정인 분산분석과 비모수 검정인 Kruskal-Wallis test를 혼용하는 식으로 모수 검정 방법과 비모수 검정 방법을 혼용한 경우가 있었다. 상관분석 시에도 특별한 언급 없이 모수 검정인 Pearson correlation test와 비모수검정인 Spearman correlation test를 혼용한 경우가 있었다.

또한 표본수가 작으면 모집단의 분포를 정확히 알기 힘들고 더욱이 정규분포 증명이 어려울 것이라는 생각으로 비모수 검정을 하는 것은 오해이다. 표본수가 작다는 점이 반드시 모집단이 비정규분포라는 증거가 될 수는 없다. t 검정이 소표본 자료에 사용되는 방법이라는 것을 감안한다면 표본의 크기가 모수와 비모수 검정을 구분하는 기준은 아니다.^{4,12,21,22} 그러나 정규성 언급 없이 단지 $N < 30$ 이므로 모수 검정인 t -test대신 비모수 검정인 Mann-Whitney U test를 시행했다고 한 경우도 있었고, 정규성 언급 없이 모수적 통계분석의 사용은 단지 표본크기가 30 이상이 되어야한다고 서술한 경우도 있었다. 또한 각 표본에서의 관찰이 다른 표본에서의 관찰과 독립적이어야 하나, 한 대상에서 수회의 관찰이나, 쌍을 이룬 관찰, 그리고 상응하는 종속 표본의 검정에서 마치 독립된 관찰처럼 취급하여 처리하면 안 된다.² 그런데도 짝을 이룬 자료나 동일집단의 반복측정인데 Paired t -test가 아닌 student's t -test를 적용한 경우가 있었다. 그 외에 변수간의 교호작용 혹은 상호작용을 알아보는 이원 분산분석 (two-way ANOVA)을 시행하고도 교호작용 혹은 상호작

용에 관한 언급조차 없는 논문도 있었다.

이러한 부적합한 분석방법의 오류가 Table VII에서처럼 2000년부터 2006년까지의 대한치과보철학회지의 한글판에서는 통계적 기법을 사용한 총 212편의 논문 중에서 47편 (22%)에, 영문판에서는 통계적 기법을 사용한 총 80편의 논문 중에서 15편 (19%)에 있었다. Table IV에서처럼 통계방법의 부적절한 기술의 오류가 발견된 128편의 논문 중에서는 한글판과 영문판을 합쳐서 62편 (48%)이 이에 해당하였다.

(3) 중복 시행한 검정

세 집단 이상에서의 평균값의 차이를 비교할 때는 처음부터 t -test를 중복 적용하면 실제론 차이가 없음에도 차이가 있는 것으로 결론을 내리게 되는 제 1종 오류 (Type I error)가 커지게 된다. 따라서 반드시 ANOVA test를 적용하여 집단 간 유의한 차이가 있는지 분석한 후에 다중비교에 의해 구체적인 차이가 있는 집단을 구분해 내야 한다. 박^{29,30}에 의하면 다중비교 방법의 선택은 다중비교를 하고자 하는 시점이나 구체적인 비교 형태 및 각 집단의 표본수가 서로 같은지가 기준이 되어 선택되어져야 한다고 했으며, 이³¹에 의하면 다중비교의 선택은 제 1종 오류를 어떻게 다루고 통제할 것인지에 따라 기준이 달라질 수 있다고 하였다.

대한치과보철학회지 논문 중에서는 동일 논문에서 여러 가지 다중비교 방법을 사용한 경우도 모두 포함시켜 살펴보니, 불분명하게 언급된 것을 포함한 전체 다중비교 방법의 사용 빈도는 총 183건 이었다. 그 중 Duncan's test가 57건 (31%), Scheffé's test가 52건 (28%), Tukey's test가 51건 (28%)으로 주로 사용되어졌다. 그 외에는 Bonferroni's test가 7건 (4%), LSD test가 6건 (3%), Student-Newman-Keuls (SNK) test가 4건 (2%) 등이 사용되어졌다. 하지만 대부분의 다중비교 방법의 선택에서 앞서 언급한 기준을 참고한 듯 보이진 않았으며, 오히려 분산분석 후 다중비교 시 특별한 언급 없이 Tukey의 방법과 Duncan의 방법을, 혹은 Tukey의 방법과 Bonferroni의 방법을, 혹은 Scheffé의 방법과 Tukey의 방법을 모두 사용하였다고 하여 유의성 있는 결과가 나오는 방법만을 선별한 건 아닌지 우려되는 오류가 있었다. 설사 두 가지 이상의 검정을 사용하여 그 결과를 비교하고 연구자가 올바르게 취사선택하였다 하더라도 최종 선택한 방법을 표기하는 것이 바람직하다고 사료된다. 그리고 대조군과 나머지 군과의 비교인 다중비교에서도 Dunnett의 방법이 아닌 Duncan의 방법을 사용한 경우도 있었다. 또한 비모수 다중비교 시에도, 즉 비모수 방법인 Kruskal-Wallis test 후 다

중비교 자체가 없거나 오히려 모수 다중비교 방법인 SNK 방법이나 Tukey의 방법을 사용한 경우가 있었고, Kruskal-Wallis test 후 비모수 방법인 Mann-Whitney U test 로 다중비교 시에도 중복가설 검정의 제 1종 오류를 줄이기 위한 P 값의 Bonferroni 수정에 관한 언급이 없었다.

이러한 중복 시행 검정의 오류가 Table VII에서처럼 2000년부터 2006년까지의 대한치과보철학회지의 한글판에서는 통계적 기법을 사용한 총 212편의 논문 중에서 34편 (16%)이, 영문판에서는 통계적 기법을 사용한 총 80편의 논문 중에서 9편 (11%)이 있었다. Table IV에서처럼 통계 방법의 부적절한 기술의 오류가 발견된 128편의 논문 중에서는 한글판과 영문판을 합쳐서 43편 (34%)이 이에 해당하였다.

3) 통계용어의 오용

국제의학학술지 편집위원회의 "생의학 학술지에 투고하는 원고의 통일양식의 통계학적 방법론에 관한 서술지침"에서도 언급했듯이, "무작위" (random) (무작위화를 의미하는 말), "정상" (normal), "유의한" (significant), "상관성" (correlations), "표본" (sample)과 같은 기술적 전문 통계용어를 비전문적으로 사용하지 않아야 하며, 통계용어, 약어, 기호의 의미를 정의해야 한다. 따라서 "유의수준 α 는 0.05 혹은 5%로 하였다"라던가 " P 값이 0.05 혹은 5% 이하인 경우 통계적으로 유의하다고 판정하였다" 로 통계학적 원리를 함축하고 있는 용어들을 그 의미의 적용이 명확하게 표현되어야 한다. 그런데도 다수의 논문에서 유의도 95%라고 표현하거나 95% 수준에서 통계학적 유의성을 검증하였다거나, 95% 유의수준으로 정하였다고 표현되어 있었으며, 유의수준 $\alpha = .05$ 가 아닌 $P = .05$ 수준으로 표기하거나, 심지어 $\alpha < .05$ 로 표현한 경우도 있었다. 이는 단순히 통계용어의 오용을 넘어 통계적 방법의 부실한 기술에 속한다고 볼 수도 있을 것이다. 또한, 유의수준이란 통계검정을 하기 전에 연구자의 의사결정 기준을 설정하는 것이므로 그 기준은 논문 전체에서 일관성이 있어야 함에도 연구방법 서술에는 유의한 차이를 $P < .05$ 인 것으로 제한하고서 정작 연구결과의 표나 내용 중에 $P < .01$, $P < .05$ 두 가지 이상의 유의수준으로 판정하는 경우가 있었으며, 심지어 유의수준 10%, 5%, 2.5%, 1% 모두에서 언급한 경우도 있었다. 그 외에 일원 분산분석인 one-way ANOVA로 기술하고서 일차원 공분산 분석이라고 언급한 경우도 있었고, 등분산 분석을 하는 Levene's test로 사후검정을 했다고 하거나, 분산분석 후에 사후 검정하는 Duncan multiple range test로 사후 분산 분석을 했다고 서술하거나, 비모수적인 t-test에 상응하는

Wilcoxon rank sum test를 비모수적인 분산분석이라 언급한 경우 등이 있었다.

이러한 통계용어의 오용에 관한 오류가 Table IV에서처럼 2000년부터 2006년까지의 대한치과보철학회지의 통계적 기법을 사용한 총 292편의 논문 중에서 통계학적 방법의 부적절한 기술의 오류가 한글판에선 114편 (54%), 영문판에선 19편 (24%)으로 전체 133편 (46%)의 논문에서 발견되었다.

고찰

2000년부터 2006년까지의 대한치과보철학회지에 게재된 논문을 크게 3가지 통계적 오류의 범주로 반복 조사한 결과, 국내외 의학학술지의 통계적 오류의 심각성을 지적한 선학들의 결과와 마찬가지로 Table IV, VII에서처럼 통계적 오류가 중복되어 나타나는 경우도 있었고, 조금만 주의를 기울이면 피할 수 있는 통계적 오류에서부터 그 내용과 의미를 이해해야 하고 연구 결과에 영향을 미칠 수 있는 통계적 오류에 이르기까지 다양하게 나타났다. 본 연구에서의 주요 결과를 검토해보면 다음과 같다.

1. 통계 프로그램의 불명시

Altman과 McDonald²³가 각종 통계프로그램을 비교분석 (benchmark analysis)한 논문에서도 언급했듯이 통계 프로그램이 복잡한 수학적 계산을 편리하게 해준 것은 사실이나 연구문제에 대한 고찰이나 자료에 대한 탐색을 소홀하게 하는 단점도 있으니 자료탐색을 소홀히 하지 말고, 모든 소프트웨어는 예외 없이 결함이 있을 수 있어 동일한 자료의 분석결과가 다소 상이하거나 오류가 나타나기도 하니, 업데이트가 가능한 프로그램의 사용을 권했으며, Bailer와 Frederick²⁴은 독자들은 자신이 사용하고자 하는 프로그램에 관해 알고 싶어 하므로 다용도의 컴퓨터 프로그램은 작동하는 컴퓨터 사양과 함께 프로그램 명칭을 명시하고 반면에 특정처리만을 위한 프로그램은 독자들도 특히 주의하고 있을 테고 자신의 연구를 위해 이용할 수 없음을 알 테니 언급하지 않아도 된다고 하였으며, 이 등⁴은 사용한 통계프로그램의 종류, 버전, 구체적인 적용기법을 기본적으로 명시하여야 하고 일부 분석기법에선 특정 프로그램이 편리하고 강력한 분석기능을 갖기도 하니 선택에 특별한 이유가 있다면 그 이유도 설명하여 후속 연구자들이 참고할 수 있도록 배려하여야 한다고 하였다.

그러나 다른 오류 항목보다 오류 빈도가 높게 나온 결

로 보아 이는 통계 프로그램 명시의 의미를 이해 못하고 간과하는 듯 보였다. 하지만 그 의미를 알아 당연하게 여기고 기술만 해준다면 본 연구에서의 다른 어떤 통계적 오류 항목보다 쉽게 개선이 가능할 것이라고 본다.

2. 통계 방법의 부적절한 기술

1) 대표값 및 산포도에 대한 부적절한 기술

자료의 특성에 따른 대표값 및 산포도에 대한 올바른 기술방법 외에도 일반적으로 통계적 검정결과를 제시하는 방법에 대한 규칙이나 관례가 있으며, 표도 자료의 성격에 따라서 전형적인 형식을 가진다. 이에 관해 APA (American Psychological Association) 출판편람 (1994)에는 심리학 분야에서 통계치를 제시하는 방법에 대한 권장 사항이 있고, 한국학술진흥재단 등재 학술지로서 A등급 판정을 받은 학술지인 "교육학연구"를 발간하는 한국교육학회 (KSSE, Korean Society for the Study of Education)에서 2007년 편찬한 "KPM (KSSE Publication Manual) 논문 작성법"에서도 통계분석 결과를 제시할 때 각 통계기법의 목적과 기능, 사용방법에 따라 유형별로 일관된 형식의 표를 작성하는 지침을 제시하고 있다. 이는 통계치를 독자적인 방법이나 단순히 통계프로그램의 결과물의 나열로 인한 혼란을 피하고 각 통계방법마다 일관되고 명확한 정보를 분명하게 전달하여 독자의 이해를 돕고자 함이다.^{4,22,25}

이렇듯 표와 도표도 논문이 주장하는 논점을 설명하고 뒷받침하는데 필요한 것에 한정시키고 독자가 알기 쉽게 정리해야 함에도 불구하고, 대한치과보철학회지의 논문에서는 그 의미를 모르는 듯 통계처리과정에서 나온 자료를 정리 없이 그대로 기술하거나 별도로 부록처럼 삽입하는 경우조차 있었다. 이는 물론 연구자의 노력도 필요하겠지만 여타 국내외 학회지처럼 비통계전문가도 쉽게 이용할 수 있도록 대한치과보철학회에서도 그에 관한 세부지침을 마련해 주는 것이 바람직할 것으로 사료된다.

2) 부적합한 분석방법

대부분의 통계처리기법은 자료의 분포에 관한 가정을 하고 있다. 가장 일반적인 형태는 자료의 정규성 (normality) 가정이며 이를 확인 하는 방법은 그림으로 자료의 빈도분포를 보거나 normal plot (정규확률 그림)을 그려보는 것이다. 이는 표본수가 적은 경우에는 효과적이나 표본수가 많은 경우 통계용 컴퓨터 프로그램을 이용하여 평가하는 방법으론 Kolmogorov-Smirnov test와

Shapiro-Wilk test가 있다. 둘째, 집단내 관찰 값들의 변동 크기가 서로 같다는 즉, 분산이 같다는 가정이 필요하다. 두개의 집단들 간의 등분산성을 검정하기 위한 방법으론 Levene's test가 있으며 세 집단 이상의 분산을 비교하기 위한 방법으론 Bartlett's test가 있다. 또한 단순회귀나 다중회귀, 변량효과모형의 경우엔 잔차 (residual)들의 분산이 서로 동질적이라는 가정도 필요하다. 셋째, 상관분석, 회귀분석의 경우에는 변수들 간의 관계가 선형적이라는 가정이 필요하다. 만약 위의 가정이 만족되지 않는다면 분석에서 요구되는 가정이 만족되도록 원자료를 적절히 변환하거나, 정규성과 같은 자료의 분포에 대한 가정을 하지 않는 비모수 검정 (非母數 檢定, nonparametric test)을 해야 한다.²¹ 모수 검정과 달리 비모수검정은 모집단에 대한 확률분포를 가정하지 않으며, 따라서 엄밀한 의미에선 다르지만 분포무관 검정 (分布無關 檢定, distribution-free test)이라고 부르기도 한다. 여기에서 '분포무관'이란 의미는 통계량의 확률분포가 모집단의 분포와 무관하다는 뜻이지, 결코 통계량의 확률분포를 가정하지 않는다는 것은 아니다. 비모수검정도 역시 관찰값이 독립적이라거나 특정 확률분포는 아니어도 모집단의 분포가 연속적이어야 한다든가 하는 가정 등이 있으나, 주로 수집된 자료가 순위척도처럼 부호나 순위의 형태일 때나 모수 검정에서 요구되는 정규 분포에 대한 가정들을 만족시키지 못하고 표본수가 적은 경우에 고려된다.^{21,22,26} 그러므로 일반적으로 가설검정에 필요한 가정을 만족시키지 못할 경우에는 비모수 검정을 이용하거나 자료의 변환을 통해 가정에 근접한 형태로 바꾸어 주어야 할 것이다.^{12,21} 하지만 대한치과보철학회지에서는 통계처리기법의 사용 전에 자료의 분포에 관한 가정에 대한 언급조차 없는 논문들이 대다수여서 이는 당연히 그런 가정의 검토를 거쳤다는 건지 아니면 검토 없이 바로 통계분석에 들어갔다는 건지 판단하기는 어려웠으며, 오히려 등분산성 검정을 위한 Levene's test로 분산분석 후에 사후 검증하였다는 식으로 기술된 걸로 볼 때 이는 그 사용의 의미는 모른 채 마치 다양한 기법을 구사한 듯이 통계처리기법을 단순히 나열한 듯한 인상을 주었다.

자료의 특성에 맞는 올바른 통계처리기법을 적용하기 위해서는 자료에 대한 객관적 탐색과정을 거쳐야 하며 이러한 절차 없이 연구자의 주관적 경험 또는 임의적 선택에 의해 적용할 통계처리기법을 선택한다면 통계적 오류는 피할 수 없다. 그러므로 통계학적 분석을 시행하기 전에는 반드시 자료의 정규성 등을 검토하여 통계처리기법을 결정해야 할 것이다.² 따라서 각 통계처리기법

에 해당하는 가정의 점검은 당연하나 모든 기법이 어렵다면 본 연구결과 Table II, III에서 보듯 대한치과보철학회지에서 자주 사용하는 통계처리기법 위주라도 최소한 알아두어야 할 것으로 보인다. 우선은 이를 위해 2005년 식품의약품안전청의 연구결과보고서 중 영국 런던 세인트조지병원에서 발간한 "Statistics Guide for Research Grant Applicants (연구 보조금 지원자를 위한 통계 지침)"을 기반으로 하여 서울대학교병원에서 주관하여 연구하고, 비통계전문가를 위해 연구계획서 작성 시 실용적인 지침서가 되도록 개발한 "임상시험 설계와 분석을 위한 기본 통계"²⁷를 숙지하고, 국제적 규모의 임상 연구자, 통계학자, 역학자, 의학 학술지 편집자들이 참여하고 많은 의학 및 보건 분야 학술지와 ICMJE (International Committee of Medical Journal Editors, Vancouver group으로 알려져 있다), CSE (Council of Science Editors), WAME (World Association of Medical Editors)와 같은 편집인 협회의 지지를 얻은 무작위 대조군 연구의 보고 방식 향상을 위한 수정 권고안인 CONSORT (Consolidated Standards of Reporting Trials, 연구 보고의 통합된 표준)양식²⁸을 참고하는 것이 도움이 될 것이다.

3) 중복 시행한 검정

모든 통계적 가설검정이 그렇듯 연구자는 자신이 밝히고자하는 가설에 대해서만 검정을 하면 된다. 대조군과 나머지 군들 간의 차이에만 관심이 있는 경우, Tukey나 Scheffé의 다중비교를 하게 되면 모든 군들 간의 차이를 비교하는 방법으로 필요 없는 검정을 추가로 하게 되어 중복가설 검정의 오류로 인해 지나치게 많이 조정된 P값을 얻게 되므로 이때는 Dunnnett의 다중비교를 하는 것이 바람직하다.²⁹ 이러한 다중비교 시에는 몇 가지 기준에 의해 다중비교 방법을 선택한다. 박^{29,30}에 의하면 모수의 다중비교에서 간단한 선택지침은 표본수가 같을 때는 Tukey, Newman-Keuls, Duncan의 방법을, 표본수가 다를 때는 Scheffé의 방법을, 대조군과 나머지군과의 비교를 위해서는 Dunnnett의 방법을 선택하라고 하였다. 또한 비모수의 다중비교에서 다중비교 방법의 선택은 독립집단에 대한 비모수 다중비교방법에는 Kraft와 Eeden, Mcdonald와 Thompeon, Miller 등의 방법이 있으나 이 검정법들은 특별한 계산표 (table)가 요구되므로 표준정규분포를 이용할 수 있는 방법으로 Dunn의 방법을 사용하라고 하였다. 그리고 관련된 집단에 대한 비모수 다중비교방법에는 모든 처리간의 비교 시에는 Harter의 방법을, 하나의 대조군과 나머지 처리간의 비교시에는 Dunnnett의 방법을 사용하라고 하였다. 아울러 이런 기준이 항상 유효

은 것은 아니나 대부분의 논문심사 과정에서는 이 기준을 지키려는 경향이 있으므로 연구자의 입장에서는 가능한 위의 기준을 따르는 것이 바람직할 것이라고까지 하였다.³⁰ 한편 이³¹에 의하면 SPSS for Windows에서의 Tukey 검정은 표본수가 같을 경우엔 Tukey HSD (Honestly Significant Differences) 검정을, 표본수가 같지 않을 경우에는 조화평균을 이용한 Tukey-Kramer 검정의 결과를 제공해 주므로 표본수가 달라도 큰 문제없다고 했다. 또한 SPSS for Windows에서는 등분산성 검정을 먼저 하여 등분산성 가정을 위반하지 않는 경우에는 평균이 유사한 동질적인 그룹으로 분류하는 범위검정 (範圍檢定, range tests)만이 필요한 경우에는 Tukey 검정을, 모든 그룹의 평균을 쌍별로 비교하여 서로 다른지 확인하는 과정인 쌍별 다중비교 (雙別 多重比較, pairwise multiple comparisons)만 필요한 경우에는 Bonferroni 검정이나 Sidák 검정을, 쌍별 다중비교와 동질적인 그룹의 분류가 모두 연구목적인 경우에는 Gabriel 검정을 추천했다. 그리고 등분산성 가정을 위반하는 경우에는 표본의 수가 50개 미만인 경우에는 Dunnnett의 T3 검정을, 50개 이상인 경우에는 GH (Games-Howell) 검정이나 Dunnnett의 C 검정을 선택하라고 권하고 있으며 이를 확대하여 표본수가 같거나 달라도, 또는 등분산성 여부와 상관없이 적용하기를 권하고 있다. 이처럼 통계전문가에 따라, 또한 통계용 컴퓨터 프로그램에 따라 약간 다른 기준을 권하고 있다.

하지만 대한치과보철학회지에서 사용된 다중비교 방법은 전체 논문에서 어떤 기준으로 선택한 다중비교 방법인지에 대한 언급조차 없어서 최소한의 올바른 기준으로 선택된 것인지 여부는 판단하기 어려웠다. 심지어 분산분석 후 다중비교에 대한 언급 자체가 없거나 분산분석의 언급 없이 다중비교의 방법만 언급한 경우도 있었다. 따라서 향후 치의학 논문 작성 시엔 사용하고자 하는 통계용 컴퓨터 프로그램에서 적절한 다중비교 방법이 선택되도록 실험계획 단계에서부터 통계 전문가와 상담이 필요할 것으로 사료된다.

3. 통계용어의 오용

의외로 다수를 차지하였으며 이는 통계용어의 정확한 정의나 이해 없이 사용하는 문제와 더불어 일견 통계서적에 따라 동일한 용어를 다양하게 표현하고 있고, 특히나 한글로 번역 시에 통일된 용어를 사용하지 않은 탓으로 인한 문제로도 보인다. 하지만 동일한 통계방법에 다양한 통계용어를 사용하면 이는 마치 다른 통계기법인 듯 착각하게 하거나 혼동하게 할 우려가 있다. 실제로 대

한치과보철학회지 논문 중에서도 student's *t*-test를 unpaired *t*-test나 independent *t*-test 혹은 two-sample *t*-test로 기술한 경우가 있었다. 따라서 통계용어의 일관되고 통일된 사용은 필요하다고 본다. 국내 통계학회에선 1987년 한국통계학회에서 편찬한 "통계용어사전"을 기준으로 하고 있으므로 가능한 이 용어사전에 있는 표현을 사용해야 할 것이다.⁴ 또한 한국통계학회 홈페이지의 자료실에서 통계학 용어 온라인 대조표를 활용하는 것도 한 방법이며, 이곳에는 1997년 한국통계학회에서 발간한 "통계학용어집 (자유아카데미 간)"의 자료 및 "통계학 용어 개정 및 온라인 서비스 위원회"에서 새롭게 추가한 내용을 웹 서비스하고 있고, 통계학 용어 대조표를 컴퓨터에 직접 다운로드 받아 설치해 사용할 수도 있으니 이를 활용하면 도움이 될 것이다.

한편 본 연구에서 사용한 통계적 오류 점검표는 통계 전문가들의 통계적 오류와 관련된 논문에서 공통적으로 자주 언급되었고 통계전문가들의 점검표 중에서 비통계 전문가도 발견할 수 있는 통계적 오류 위주로 항목을 구성하였으나, 평가항목을 나열식으로 제시했을 뿐 개별 항목의 오류에 대한 경중을 구분하지는 않고 있기 때문에 점검표에 의한 평가결과를 해석할 때 오류를 범한 항목의 수에만 의존하여 그 논문에 대한 통계적 방법론의 타당성을 평가하는 것은 무리가 있을 수 있다.²⁴ 따라서 전체적인 통계적 오류의 빈도 그 자체보다는 오류를 범한 내용이 의학적 결론에 미치는 영향에 관심을 가져야 할 것이며, 통계적 방법 자체도 중요하지만 통계학적 방법론에 대한 인식을 제고하는데 우선을 두어야 할 것으로 사료된다. 본 연구 중에도 여타 몇몇 연구들처럼^{32,34} 그러한 통계적 오류의 항목 수의 나열만 되는 상황을 최대한 피해보고자, 점검표를 이용한 반복 평가 시에도 오류 건수의 나열이 아닌 논문의 편수로 표현하려 시도하였으며 무리하게 타당도 점수로 평가하지 않았다. 또한 언급이나 기술이 안 되었다고 무조건 통계적 오류로 단정하지 않고 가능한 한 기술되어 있는 것 중에서 명확한 통계적 오류로 판단되는 것을 위주로 반복 평가하였다. 하지만 향후 비통계전문가 입장에서 좀 더 용이하게 활용할 수 있는 점검표와 그에 관한 검토가 필요하다고 본다.

본 연구의 결과에서 보듯이 향후 치의학 논문에서 통계적 방법을 기술하고 적용할 때는 연구자 본인이 실제 통계적 분석을 수행하지는 않았더라도 그 내용과 결과에 대해선 연구자로서의 책임이 있으므로 통계적 오류를 피하기 위해 주의를 기울여야 할 것으로 보이며, 또한 선학들의 지적처럼⁴⁶ 가능하다면 실험계획 단계에서부터 통계 전문가와 상담하여 바람직한 실험설계 및 적절한

한 통계적 방법의 적용이 가능토록 해야 할 것으로 생각된다. 아울러 치의학 연구 논문에서의 통계학적 방법론에 대한 인식을 제고하여 적절한 통계적 방법의 적용과 지속적인 통계적 오류 개선의 노력이 있어야 할 것으로 사료된다.

결론

대한치과보철학회지의 통계적 오류현황을 파악하고자 2000년부터 2006년까지의 대한치과보철학회지의 총 399편의 논문 중 통계적 방법을 사용한 전체 292편의 논문에서 한글판 212편과 영문판 80편의 통계적 오류 항목을 크게 3가지 범주로 나누어 비교 조사해 본 결과 전체적으로 대한치과보철학회지에 게재된 논문의 통계적 타당성은 다소 부족한 걸로 나타났다. 본 연구에서의 주요 결론은 다음과 같다.

1. 사용된 통계 프로그램의 종류는 SPSS (73%), SAS (20%), Microsoft Excel (4%), Statgraphics plus (1%), SigmaStat for Windows (1%), Minitab (1%) 순으로 많이 사용되었으며, 통계 프로그램 불명시의 오류는 한글판에선 54%가, 영문판에선 59%가 있었으나 통계적으로 유의한 차이를 보이지 않았다 ($P = .66$).
2. 대표값 및 산포도에 대한 부적절한 기술, 부적합한 분석 방법 및 중복 시행한 검정의 오류 등을 포함한 통계 방법에 대한 부적절한 기술에 관한 오류는 한글판에선 49%가, 영문판에선 33%가 있었으나 통계적으로 유의한 차이를 보이지 않았다 ($P = .09$).
3. 기술적 전문 통계용어를 비전문적으로 사용하거나, 통계용어, 약어, 기호의 의미를 잘못 정의하는 통계용어의 오용 오류는 한글판에선 54%가, 영문판에선 24%가 있었으며 통계적으로 유의한 차이가 있었다 ($P = .01$).
4. 상기와 같이 대한치과보철학회지의 통계적 오류현황은 통계 프로그램 불명시의 오류와 통계적 방법에 대한 부적절한 기술에 관한 오류 및 통계용어의 오용 오류에서 한글논문은 49% - 54%, 영문논문은 24% - 59%가 나타났다. 통계용어의 오용 오류 항목 이외엔 통계적으로 유의성이 있진 않았으므로 한글논문과 영문논문의 통계적 오류의 비율은 유사하게 나타난 걸로 추정된다.

따라서 대한치과보철학회지에 투고하는 치의학 연구자는 실험계획 단계에서부터 통계 전문가와의 상담 및 치의학 연구 논문에서의 통계학적 방법론에 대한 인식을 제고하여 적절한 통계적 방법의 적용과 통계적 오

류 개선의 노력이 있어야 할 것이며, 아울러 대한치과보철학회에서도 치의학 연구자가 각 통계적 방법마다 일관되고 명확한 정보를 분명하게 전달하여 독자의 이해를 돕도록 여타 국내외 학회지처럼 비통계 전문가도 쉽게 이용할 수 있는 세부지침을 마련해 주는 것이 바람직할 것으로 사료된다.

참고문헌

- Choi JH, Lee JC. Scholarly Monographs and Statistical Methods. Seoul: Freedom Academy Pub. Co.; 1990.
- Lee HJ, Kim DH. Research Methodology Topics in Statistics. Seoul: Korea Medical Book Publisher; 2002.
- Altman DG. Statistical Reviewing for Medical Journals. *Stat Med* 1998;17:2661-74.
- Lee HJ. Medical Research and the Use of Statistical Methods. Seoul: Korea Medical Book Publisher; 2000.
- Choi JH, Kim HG. Scientific Methodology and the Use of Statistical Methods. Seoul: Freedom Academy Pub. Co.; 1994.
- White SJ. Statistical Errors in Papers in the British Journal of Psychiatry. *Br J Psychiatry* 1979;135:336-42.
- ICMJE(International Committee of Medical Journal Editors). Uniform Requirements for Manuscripts submitted to Biomedical Journals. *J Am Osteopath Assoc* 2003;103:137-49.
- Ko UR, Ahn YO. Statistical Resumption for Health Articles in One Series of Medical Journals. *Korean J Prev Med* 1973;6:81-6.
- Ko H. Statistical Methods in the Articles of the Journal of the Korean Society of Anesthesiologists from 1981 to 1990. *Korean J Anesthesiol* 1993;26:22-7.
- Kim YY, Lee EI, Jung HR. Research Designs, Research Methods, and Statistical Techniques in Korean Ophthalmic Articles. *J Korean Ophthalmol Soc* 1994;35:460-73.
- Hwang KH, Lee HJ, Kim YJ, Lee SI. Statistical Errors in Papers in the Journal of Korean Society of Plastic and Reconstructive Surgeons. *J Korean Soc Plast Reconstr Surg* 2001;28:302-9.
- Publication Committee of Korean Academy of Family Medicine. Statistics for Family Medicine. Seoul: J Korean Acad Fam Med; 2001.
- Koh WJ, Lee SJ, Kang MJ, Lee HJ. An Evaluation of the Use of Statistical Methods in the Journal of Tuberculosis and Respiratory Diseases. *Korean Tuberc Respir Dis* 2004;57:168-79.
- Pak SI. An Assessment of Statistical Errors in Articles in the Korean Journal of Veterinary Research. *Korean J Vet Res* 1999;39:1187-96.
- Lee SD. An Assessment of Statistical Validity of Articles Published in the Journal of Korean Acupuncture and Moxibustion Society from 1984 to 2002. *Korean J Acupunct Mox* 2004;21:176-88.
- Park YG. The Most Common Statistical Errors in the Articles. *J Korean Acad Fam Med* 1998;19:137-43.
- Gore SM, Jones JG, Rytter EC. Misuse of Statistical Methods: Critical Assessment of Articles in *BMJ* from January to March 1976. *Br Med J* 1977;1:85-7.
- Ahn YO, Lee HK. Development of a Checklist for Assessing the Methodological and Statistical Validity of Medical Articles. *Korean J Med Educ* 1991;3:19-35.
- Park YG. Statistical Application of the Articles in the Journal of the Korean Academy of Family Medicine. *J Korean Acad Fam Med* 1998;19:137-43.
- Lee HK, Huh BY, Ahn YO. An Assessment of Methodological and Statistical Validity of Medical Articles Published in Korea, from 1980 to 1989. *J Korean Acad Fam Med* 1991;12:46-67.
- Petrie A, Sabin C. Medical Statistics at a Glance, 2nd edition. London: Blackwell publishing Ltd; 2005.
- Ahn YO, Yoo KY, Park BJ. Applied Medical Statistics, Revised edition. Seoul: Seoul National University Press; 2005.
- Altman M, McDonald MP. Choosing Reliable Statistical Software. *Political Science and Politics* 2001;34:681-7.
- Bailar JC 3rd, Frederick M. Guidelines for Statistical Reporting in Articles for Medical Journals: Amplifications and Explanations. *Ann Intern Med* 1998;108:266-73.
- Browner WS. Publishing and Presenting Clinical Research. Philadelphia: Lippincott Williams & Wilkins Ltd; 1998.
- Song MS, Park CS. Introduction to Nonparametric Statistics. Seoul: Freedom Academy Pub. Co.; 1997.
- Hahn SK. Development of Program for Statistical Education in Design and Analysis of Clinical Trials and Training Medical Statisticians. The Annual Report of KF-DA No. GOVP1200608585, Seoul: Korea Food and Drug Administration; 2005.
- Altman DG, Schultz KF, Moher D, Egger M, Davidoff F, Elbourne D, Lang T, et al. The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration. *Ann Intern Med* 2001;134:663-94.
- Park YG. Statistical Errors of the Articles in December 1999. *J Korean Acad Fam Med* 2000a;21:107-10.
- Park YG. Statistical Errors of the Articles in January 2000. *J Korean Acad Fam Med* 2000b;21:289-93.
- Lee YJ. Understanding the Analysis of Variance. Seoul: SukJung publisher; 2000.
- Park GB. A Study of Statistical Analysis Methods in Domestic Journal related to Oral and Maxillofacial Surgery. MSD. Graduate School of Chosun University; 1999.
- Lee SM, Lee SW. An Assessment of Statistical Validity of Nursing Researches Published in Korea from 1986 to 1995. *J Korean Soc Health Statistics* 1998;23:42-64.
- Park IK, Kang JO, Kim TY, Keum DG. An Evaluation of the Statistical Techniques Used in the 1995-1996 Editions of the Korean Journal of Clinical Pathology. *Korean J Clin Pathol* 1999;19:460-4.

An assessment of statistical errors of articles in the Journal of Korean Academy of Prosthodontics: Comparison between Korean version and English version

Dong-Gyu Park¹, DDS, MSD, Yong-Geun Choi², DDS, MSD, PhD,

Young-Su Kim², DDS, MSD, PhD, Sang-Wan Shin^{2*}, DDS, MPH, PhD, MSc.

¹Graduate student, ²Professor, Department of Advanced Prosthodontics, Graduate School of Clinical Dentistry, Korea University, Korea

Statement of problem: The aim of dental research is to advance scientific knowledge and leads to improvement in the treatment and prevention of dental disease. Utilizing an effective research design and adequate statistical methods are essential procedures ensuring that the results of researches are based on evidences. A research should utilize proper statistical methods without statistical errors; Otherwise, it could adversely affect clinical practice and future research. **Purpose:** This study was made to investigate the statistical methods used in the Journal of Korean Academy of Prosthodontics (JKAP) and then to assess them for the statistical errors. **Material and methods:** Among the total of 399 articles in the JKAP published from 2000 to 2006, 292 articles using statistics were reviewed. The validity of the statistical methods used in them were assessed using a checklist based on the guideline for statistical reporting in the uniform requirements for manuscripts submitted to biomedical journals by International Committee of Medical Journal Editors. The checklist consisted of three categories of statistical errors: 1) Unspecified computer statistical packages, 2) Inadequate description of statistical methods, 3) Misuse of statistical terms. Then, the results were compared between the Korean version and the English version in the JKAP. **Results:** Among the 212 articles using statistics in the Korean version, 115 articles (54%) and among the 80 articles using statistics in the English version, 47 articles (59%) were shown to have unspecified computer statistical packages without statistically significant difference ($P = .66$). Likewise, 101 articles (48%) in the Korean version and 25 articles (31%) in the English version were shown to have the inadequate description of statistical methods without statistically significant difference ($P = .09$). However, 114 articles (54%) in the Korean version and 19 articles (24%) in the English version were shown to have the misuse of statistical terms with statistically significant difference ($P = .01$). **Conclusion:** Some of the articles in the JKAP had inadequate statistical validity, given the statistical errors identified in this assessment. Hence, dental researchers should be more careful when it comes to describing and applying statistical methods. (*J Korean Acad Prosthodont* 2009;47:273-85)

Key words: statistical error, misuse of statistical method, statistics, research design, prosthodontics

Corresponding Author: Sang-Wan Shin

Department of Advanced Prosthodontics, Graduate School of Clinical Dentistry, Korea University, 126-1 Anam-dong 5-Ga, Seongbuk-Gu, Seoul, 136-705, Republic of Korea
+82 2 818 6874, swshin@korea.ac.kr

Article history

Revised May 19, 2009 / Last Revision June 17, 2009 / Accepted July 2, 2009