

의학학술지 게재논문들에서의 통계학적 방법론 활용현황 및 소아알레르기 호흡기학회지의 질 향상을 위한 제언

인하대학교 의학전문대학원 사회의학교실

이 훈 재 · 정 수 경

=Abstract=

The Use of Statistical Methodology in Articles in Medical Journals and Suggestions for the Quality Improvement of the Pediatric Allergy and Respiratory Disease

Hun Jae Lee, MD, Su Kyung Jung, MS

*Department of Social and Preventive Medicine,
Inha University School of Medicine, Incheon, Korea*

Statistical analysis is an essential component of all biomedical research. Use of descriptive and inferential analysis enables researchers to summarize findings and conduct generalizations from research findings. Physicians must keep current with clinical information to practice evidence-based medicine. In doing so, physicians need to access reports of original research. This requires the reader to critically appraise the design, conduct, and analysis of each study and subsequently interpret the results. However, many surveys reveal prevalent statistical errors in articles in medical journals, and it is also clear that many physicians are seriously unfamiliar with statistical methods. This constitutes a major barrier to the development of medical research. Given this situation, this review article discusses the use of statistical methods in medical journal articles, the status of statistical errors among them, and the results of evaluation of physicians' understanding of statistical methods. Moreover, this article introduces the current trend in statistical reviews to improve the quality of medical journal articles, and explains the uniform requirements for manuscripts submitted to medical journals. Finally, the article proposes several specific ways to improve the quality of the Pediatric Allergy and Respiratory Disease. [Pediatr Allergy Respir Dis (Korea) 2011;21:144-155]

Key Words : Use of statistical analysis, Statistical error, Statistical review

서 론

증례보고나 종설 등을 제외한다면 대부분의 의학논문에서는 표본을 대상으로 얻은 자료를 요약하거나 그 결과를 일반화하기 위하여 통계학적 원리와 기법을 활용하고 있다.

따라서 통계학적 방법론은 의학연구에 있어서 가장 중요하고 보편적인 방법론으로 강조되고 있으며, 이의 올바른 활용 여하에 따라 의학논문의 학술적 가치와 임상적 유용성 등이 판가름 난다고도 한다.^{1,2)} 아울러 의학연구라는 것은 의학현상에 내재하고 있는 진리를 객관적 방법론에 의하여 규명하는 과정이므로, 조금 거창하게 말한다면 통계학적 방법론의 올바른 활용은 의학을 과학으로 규정짓게 하는 문제와도 직결된다고 할 수 있을 것이다. 대부분의 외국 저명 의학학술지에서는 이러한 통계학적 방법론의 중요성을 인식하여 지

접수: 2011년 9월 16일, 승인: 2011년 9월 19일
책임저자: 이훈재, 인천시 중구 신흥동 3가 정석빌딩 C동 3층
인하대학교 의학전문대학원 사회의학교실
Tel : 032)890-0974 Fax : 032)884-6725
E-mail : lee4146@inha.ac.kr

난 1960년대부터 투고논문들에 대한 별도의 통계사독을 시행하고 있기도 하다.³⁾

최근에 와서는 일반 의사들의 의학통계에 대한 이해와 활용능력이 부족 강조되고 있다. 통계는 평생의학학습에 필요한 기본 언어이기도 하며, 근거에 입각한 의료를 행하기 위한 핵심 소양이기 때문이다.⁴⁻⁶⁾ 평생 동안 직업적 전문성을 유지, 심화시켜야 할 의사들로서는 전문서적 또는 의학학술지를 의학지식 습득경로로서 선호하고 있다. 전문서적이거나 의학학술지 지면 내용의 상당부분은 통계학적 원리와 기법을 활용하여 도출된 의학연구의 결과들이다. 따라서 통계에 대한 어느 정도의 식견이 없다면 그것들을 비판적으로 탐독할 수는 없을 것이다. 질 관리가 잘되고 있는 외국 의학학술지의 경우도 통계기법을 활용하고 있는 논문의 40-70% 정도에서 통계학적 오류가 확인되고 있다고 한다. 대한의사협회지, 대한마취과학회지, 대한류마티스학회지, 대한방사선학회지, 대한성형외과학회지, 대한안과학회지, 대한이비인후과학회지 등 우리나라 주요 의학학술지 게재논문들의 경우는 대략 60-100% 정도에서 각종 오류를 범하고 있다고 한다.⁷⁾ 그러나 이를 개선하기 위한 국내 학회 또는 학술지 차원의 구체적인 활동은 활성화되지 못하고 있는 실정이다.

이에 본 고에서는 의학학술지 게재논문들에서의 통계활용 및 통계학적 오류 실태, 의사들의 통계학적 방법론 이해도 평가결과, 의학학술지 게재논문들의 질 향상을 위한 통계사독의 현황, 의학학술지 투고논문들이 준수하여야 할 통계학적 방법론 서술 및 결과제시 원칙, 소아알레르기 호흡기학회지 게재논문들에서의 통계활용 현황을 검토해보았다. 그리고 투고규정의 개정 등 소아알레르기 호흡기학회지의 질 향상을 위한 몇 가지 방안을 제언하는 것으로 결론 맺고자 하였다.

의학학술지 게재논문들에서의 통계활용 및 통계학적 오류 실태

1966년, Schor과 Karten³⁾은 Journal of the American Medical Association (JAMA)을 비롯한 10대 주요 의학학술지 게재논문 295편에서의 통계활용과 통계학적 오류 실태를 평가하여 보고하였다. 이들의 연구결과에 의하면 평가대상 논문의 28%만이 통계학적으로 수용 가능할 뿐이며 각각 67%와 5%는 통계적으로 결함이 있거나 수용자체가 불가능한 수준이었다. 이 연구는 의학학술지 게재논문들에 대한 질 관리와 통계활용의 적정성 모니터링이 필요하다는

공감대 형성의 계기가 되었다. 또한 이 연구와 유사한 연구도 이어지게 되었는데 British Medical Journal (BMJ),⁸⁾ British Journal of Psychiatry,⁹⁾ 그리고 Arthritis and Rheumatism¹⁰⁾에 게재된 논문들을 평가한 연구들이 널리 알려져 있다. 이 연구들에서는 각각 평가대상 논문의 52%, 45%, 68% 정도에서 크고 작은 통계학적 오류가 나타났다고 결론내고 있다. 세 연구 모두에서는 부적절한 대표값과 산포도 지표의 제시, Student's *t*-test와 카이제곱검정 적용상의 오류가 가장 흔한 통계학적 오류 유형이었음을 보여주고 있다.

의학학술지 게재논문 대상의 평가결과를 통해 드러난 문제점과 1965년부터 통계사독을 실시한 JAMA의 성과가 알려지면서 주요 의학학술지 대부분은 속속 통계사독 제도를 도입하기 시작하였다. 이러한 동향에 따라 국제의학학술지 편집인위원회(International Committee of Medical Journal Editors)에서는 1988년 '생의학 학술지에 투고하는 원고의 통일양식' 개정 3판을 발간하며 통계 서술지침을 명시하였다. 총 15개항으로 구성된 이 통계 서술지침에서는 주로 게재신청 원고의 작성과 관련한 통계학적 방법론 및 결과 서술요령뿐 아니라, 연구자들이 준수하여야 할 핵심적인 통계학적 방법론의 적용 원칙을 포괄하고 있기도 하다. 이 지침은 대부분 의학학술지에서 투고규정의 내용으로 채택하였으며, 게재신청 논문에 대한 통계사독의 기준이 되었다.¹¹⁾

그 결과 외국 주요 의학학술지 게재논문에서의 통계학적 오류 실태는 어느 정도 개선되는 효과를 거둔 것은 분명하지만 아직까지도 연구결과의 타당성을 훼손할 수 있는 오류들이 지속되고 있다. 이를테면 세계 최고의 의학학술지로서 통계학적 방법론의 중요성을 특히 강조하며 정기적으로 의학통계 지면강좌를 운영하고 있는 BMJ에서조차도 확률개념에 해당하는 민감도(sensitivity)를 108%라고 제시한 논문이 게재되기도 한다.^{12,13)} 이런 오류는 의학학술지 게재논문에서 지금도 간혹 볼 수 있다.

의학학술지 게재논문들의 통계학적 오류 평가가 초기에는 주로 의학통계학자들에 의해 이루어졌으나 1990년대 이후부터는 의학연구자들이 직접 수행하는 경향이 뚜렷하였다. 의학연구자들의 눈높이에서 의학학술지 게재논문들이 준수하여야 할 통계학적 방법론 활용의 적정성을 구체적으로 평가한다는 점에서 초기 실태조사 연구들에 비해 그 의의가 더욱 크다. 그 중 대표적인 것으로는 Annals of Internal Medicine, BMJ, JAMA, Lancet, New England Journal of Medicine (NEJM), Canadian Medical Association Journal 등 6종의 저명 의학학술지 게재논문의 초록에

기제된 통계서술의 타당성을 검토한 연구가 있다.¹⁴⁾ 이 연구에서는 평가대상 학술지 게재논문들의 18-68%가 본문의 내용과 다르거나 본문에 제시되지 않은 통계결과를 초록에 제시하고 있음을 보여주었다. 통계기법 적용의 오류라기 보다는 결과서술상의 실수로 간주되기도 하는 오류이다. 그러나 다수의 독자들이 논문 초록을 통해 연구의 결과와 가치를 판단하고 있다는 사실을 감안한다면 사소한 실수로 넘어갈 문제는 아닌 것이다. Obstetrics & Gynecology 등 주요 산부인과학술지 게재논문들 중 로지스틱회귀분석 활용의 타당성을 평가한 연구에서는 이 통계기법의 적용을 위해 검토하거나 지켜져야 할 원칙들이 무시되고 있다고 결론내고 있다.¹⁵⁾ 평가대상 논문 193편 중 93.2%가 로지스틱회귀분석 모형의 적합도를 검토하지 않았으며, 86.4%는 독립변수 간의 상호관련성을 분석하지 않았고, 51.8%가 독립변수 선정 원칙 등을 제시하지 않은 것에 근거한 결론이었다. 외삽(extrapolation)이라고 하는 것은 상관 및 회귀분석의 결과를 제시하거나 해석할 때 흔히 범하게 되는 중요한 오류이다. 세계 4대 의학학술지라고 할 수 있는 BMJ, JAMA, Lancet 그리고 NEJM에 게재된 논문 중 산점도(scatter plot)를 그려 핵심적인 연구결과를 요약제시하고 있는 37편의 논문을 검토한 연구에서는 전체의 60%가 외삽의 오류를 보인 것으로 평가되었다.¹⁶⁾ 한편, 중국의 5대 의학학술지 게재논문들에서의 통계활용과 통계학적 오류 실태를 평가하여 JAMA에 투고되었던 연구결과에서는 통계기법을 활용한 논문의 36.3%에서 어떤 통계기법을 적용하였는지 구체적으로 명시하지 않았으며, 53.6%에서는 부적절한 통계기법을 적용하고 있었다고 했다. 부적절한 통계기법의 적용 사례로는 analysis of variance 또는 Paired *t*-test를 적용해야 할 자료를 Student's *t*-test로 분석한 것이 가장 흔하였다.¹⁷⁾ Infection and Immunity지에 게재된 141편의 논문을 평가한 연구에서는 전체 논문의 54%에서 통계학적 오류가 있다고 하였는데, 이를 세분화하면 20%는 통계기법 적용상의 오류였고 22%는 통계결과 서술상의 실수였으며 나머지 12%는 통계기법 적용과 결과서술 모두에 결함이 있었던 것이었다.¹⁸⁾

중국의 경우만 해도 주요 의학학술지에서는 이미 1990년대부터 통계사독을 본격적으로 실시해오고 있지만,¹⁷⁾ 국내에서 발간되는 의학학술지 중 통계사독을 별도 시행하는 경우는 거의 없는 실정이다. 의학학술지의 질 관리와 향상이라는 책임이 사실상 저자와 사독위원의 개인적 역량에 전가되고 있는 상태인 것이다. 따라서 국내 의학학술지 게재논문들에서 통계학적 오류가 만연함은 어찌 보면 당연한 현상이

다. 국내에서는 1980년대 발간된 대한의학협회지 게재논문들을 대상으로 통계학적 오류 실태를 평가하여 보고²⁾한 것이 관련 연구의 발단이 되었다. 이 연구에서는 통계적 기법을 활용한 논문 297편의 97.8%인 290편에서 한 가지 이상의 통계학적 오류를 보였다고 평가하였는데, 흔한 오류 유형으로서는 통계학적 검정력과 신뢰구간에 대한 불충분한 기술(91.9%), 잘못된 통계기법 적용으로 인한 중복검정의 오류(65.2%), 통계기법 활용에 대한 불충분한 기술(58.2%), 불합리한 통계학적 결론도출(52.2%) 등 이었다. 대한안과학회지 게재논문들의 통계학적 방법론을 평가한 연구에서는 통계기법을 활용한 평가대상 89편의 논문 모두에서 한 가지 이상의 통계학적 오류가 나타났고, 32.5%는 어떤 통계기법을 적용하였는지를 명시하지 않았다고 하였다.¹⁹⁾ 대한이비인후과학회지 게재논문 중 통계기법이 활용된 153편의 논문을 대상으로 한 연구²⁰⁾에서는 전체의 36.6%만이 통계기법 활용상의 특별한 오류가 없었을 뿐 나머지 63.4%에서는 통계기법 활용상의 오류가 있거나 필요한 통계분석을 시행하지 않은 것으로 나타났다. 평가대상 논문의 31.4%는 부적절한 통계기법을 적용하였고, 19.0% 가량은 *P* 값을 토대로 결론을 도출하고는 있으나 어떤 통계기법이 적용된 것인지를 명시하지 않고 있었다.

본 저자의 경우도 일부 국내 의학학술지 게재논문들을 평가한 경험이 있다. 대한성형외과학회지에 게재된 논문 중 통계기법을 활용한 167편을 검토한 결과 전체의 47.0%에서 적절하지 않는 대표값과 산포도 지표로 자료를 요약하고 있었으며, 67.0%가 자료의 특성에 맞지 않는 통계검정 기법을 적용한 것으로 확인할 수 있었다. 국제의학학술지 편집인위원회에서 제시한 '생의학 학술지에 투고하는 원고의 통일양식'의 통계 서술지침 세부항목에 대한 평가결과에서는 게재 논문의 26.8%만이 통계학적 방법론과 분석결과를 비교적 상세히 기술하고 있었으며, 유의수준 또는 신뢰구간을 제시한 경우는 전체의 37.5% 그리고 *P* 값을 정량적으로 기술하고 있는 논문은 25.3%였다. 또한 연구방법에서 적용하였다고 소개한 통계기법의 결과를 모두 제시한 경우는 61.2%였으며, 19.2%에서만이 활용한 통계 프로그램의 종류를 명시하고 있었다. 전체 평가대상 논문의 47.9%에서 통계용어를 올바르게 사용하고 있을 뿐이었다. 한편, 평가대상 논문 모두에서 통계학적 방법론 활용을 위한 구체적인 참고문헌을 제시하지는 않고 있었다. 이러한 결과를 종합할 때 통계검정 기법을 활용한 논문의 5.5%만이 통계사독을 무난히 통과할 수 있을 것으로 판단되었다.²¹⁾

결핵 및 호흡기질환 게재논문 중 통계기법을 활용한 107

편의 논문을 평가하였던 연구에서는 83.2%에서 적절하지 못한 대표값 또는 산포도를 제시한 것으로 나타났다. 통계검정 등 추측통계학적 분석기법을 활용한 논문 89편 중 79편(88.8%)은 자료의 특성에 맞지 않는 잘못된 통계기법을 적용하고 있었으며, 가장 흔한 오류는 정규성이 의심되는 변수에 모수검정기법을 적용한 경우였다.(35편, 39.3%) 또한 연구방법을 기술하며 적용하였다고 소개하였던 통계기법들의 분석결과를 누락한 경우도 31편(34.8%)이 있었다. 전체적으로 통계기법이 활용된 논문의 86.9% 가량이 연구결과의 타당성을 크게 훼손할 수 있을 정도의 통계학적 오류를 범한 것으로 평가되었다.⁷⁾

의사들의 통계학적 방법론 이해도 평가결과

의학논문에서 활용되고 있는 다양한 연구방법론 중에서 보편성과 연구결론 도출과의 밀접성이 가장 큰 것은 바로 통계학적 방법론이다. 그러므로 통계의 기본개념을 이해하지 못하고서는 의학논문을 비판적으로 읽는 것이 불가능하다는 것이다. 즉, 통계는 의학연구를 위한 핵심적인 방법론이자 의학지식 습득을 위한 기본 언어라 할 수 있다. 국내 의학분야에서도 통계학적 방법론의 올바른 활용에 대한 관심이 커져가고 있으며, 의과대학과 의료기관 그리고 학회 차원에서의 통계강좌도 활성화되고 있는 추세이다. 그러나 이와 관련한 현황을 면밀히 파악하지는 못하고 있으며, 의학통계 교육의 내실화 방안이 체계적으로 검토되지도 못하고 있는 실정이다. 그러나 외국의 경우 우리에게도 큰 시사점을 주는 다양한 현황과 노력의 시도가 있다.

지난 1981년 Pediatrics지를 통해 보고된 Friedman과 Phillips의 연구²²⁾가 대표적인 예이다. 이 연구에서는 미국 주요 종합병원 소아과 전공의 684명을 대상으로 핵심적인 의학통계용어 두 가지 즉, 상관계수와 유의수준에 대한 이해도를 평가하였다. 상관계수는 의학연구에서 흔히 활용되는 대표적인 통계기법의 결과지표이며, 유의수준은 통계검정의 핵심원리를 함축하고 있는 용어이다. 따라서 이 두 용어에 대한 이해를 못하고 있다면 의학학술지 게재논문을 제대로 읽지 못할 것이라고 저자들은 가정을 한 것이다. 연구결과 미국 소아과 전공의 1, 2, 3년차의 각각 17.3%, 16.2%, 15.8%만이 상관계수의 크기를 올바르게 판단하는 것으로 나타났다. 유의수준에 대한 개념을 묻는 문항의 정답률은 각각 55.5%, 61.3%, 49.2%였다. 정답률이 예상보다 낮기도 하였지만, 두 문항 모두 의학논문을 접할 기회가 많았던 3년차 전공의들에서의 정답률이 상대적으로 낮다는 것도

문제였다. 이러한 결과를 근거로 미국 소아과 전공의들의 상당수가 의학학술지 게재논문을 제대로 읽는 것이 어려운 상태라고 결론을 내었으며, 의과대학과 전공의 수련기관 차원에서의 통계교육 프로그램의 활성화가 제언되었다.

미국의 의과대학생, 전공의, 대학교수, 개업의 281명을 대상으로 통계의 기본개념에 대한 이해도와 간단한 확률계산 능력 등을 평가한 연구에서도 비슷한 결과를 보여주고 있다.²³⁾ 이 연구에서는 의학통계의 핵심적인 개념들을 반영한 총 5개 영역 41개 세부문항으로 구성된 평가시험도구를 활용하여 대상자들의 의학통계에 대한 이해를 평가한 것이다. 일종의 객관식 평가시험 결과를 100점 만점으로 환산을 하였을 때 전체 응답자들의 평균점수는 63점이었다. 소집단별로 비교하면 의과대학생, 전공의, 개업의, 대학교수들의 점수는 각각 평균 73점, 70점, 55점, 74점이었다. 대학교수의 의학통계에 대한 이해도가 의과대학생과 유사할 뿐, 전공의와 개업의들의 수준은 의과대학생에 비하여 낮았다. 특히 개업의들의 경우 통계계산 등 일부 영역에서는 의학통계에 대한 지식이 전무한 상태에서 무작위로 응답할 때 얻을 것으로 기대되는 수준의 평균점수를 보였다. 통계에 대한 중앙감으로 인하여 알고 있는 지식마저도 제대로 활용하지 못함을 시사하는 결과이다.

한편, 의과대학 졸업 후 5년이 경과된 외과 전공의들을 대상으로 조사한 결과에서는 90% 정도가 전공의 훈련과정 중 체계적인 의학통계 교육이 필요하다고 응답하였다. 그러나 전공의 훈련과정 중 의학통계에 대한 교육을 받을 기회가 총 5시간 미만이었다고 답한 경우가 전체의 92%나 되었다.²⁴⁾ 이러한 결과들은 각급 의학교육 기관들로 하여금 다양한 통계교육 프로그램의 운영을 촉진하였다.^{5,25,26)}

미국 전공의 277명을 대상으로 통계에 대한 지식과 태도 등을 평가한 연구논문이 비교적 최근에 JAMA에 게재되었다. 이 논문에서는 미국 전공의들의 의학통계 개념에 대한 이해수준이 100점 만점에 평균 41.4점에 그치고 있음을 보여주며, 전공의 대다수가 의학논문을 비판적으로 읽는 것이 곤란하다는 문제를 다시 제기하고 있었다. 그러면서 전공의 수련과정에 의학통계 등에 대한 교육이 더욱 강화되어야 함을 역설하고 있다. 통계에 대한 전공의들의 평균적인 이해수준이 낮기는 하였지만, 대학 또는 전공의 과정 중 의학통계 교육의 기회가 있었던 응답자들에서 그 이해도가 높은 수준이었다. 이는 의과대학이나 수련기관 차원에서의 의학통계 교육의 필요성을 의미하는 것이다.

의학학술지 게재논문들의 질 향상을 위한 통계사독의 현황

통계학적 방법론 활용의 적정성을 평가하는 통계사독은 전체적인 사독 활동에서 차지하는 비중이 매우 크다. Medline 검색엔진으로 사독의 영문용어에 해당하는 ‘peer review’를 검색해 보면 대략 12,000여건의 문헌자료를 얻을 수 있다. 이들 문헌의 대부분은 주로 편집인의 사설(editorial), 독자의 논평 또는 편지(commentary or letters) 형식이지는 않지만, 사독 과정의 문제점과 개선방안을 주제로 한 연구논문도 수백 건 이상이 포함되어 있는 듯하다. 의학 논문의 사독 자체를 탐구한 문헌들이 예상외로 많다는 점이 놀랍기도 하며, 그 대부분이 1980년대 후반 이후 집중 발표되고 있음을 고려하면 최근 국제 의학계에서 일고 있는 사독에 대한 관심과 고민이 얼마나 큰지를 짐작할 수 있을 것이다. 우리나라의 경우 아직 일부 학술지 편집인과 학회회원 등이 중심이 되어 사독 제도의 개선을 위한 논의가 간헐적이고 비공식적으로 진행되고 있는 수준이다. 국내 의학연구가 적어도 양적인 측면에서는 비약적인 성장이 있었으나 질적인 발전은 그에 미치지 못하고 있다는 지적에 대다수 국내 의학자들은 공감할 것이다. 의학연구의 발전을 위해서는 관련 제도의 뒷받침이 필요하지만 각 학회 차원에서 노력해야 할 일들도 많다. 학회 차원에서 해야 할 노력 중에서는 발간학술지 게재논문의 질 향상을 위한 사독제도 운영의 내실화가 특히 중요하다.

사독의 사전적 정의는 “의학현상 탐구를 위한 연구자의 아이디어, 계획서, 결과물을 동료연구자들이 평가·검증하는 과정”이라고 할 수 있다. 따라서 의학학술지 게재신청 논문, 학술대회 발표신청 초록, 연구비 신청을 위한 계획서에 대한 평가는 물론 이미 발표된 연구논문 등에 대한 동료 의학자 등의 건설적 비판도 넓은 의미의 사독 범주에 포함된다. 전자와 후자를 각각 사전심사(pre-publication review)와 사후심사(post-publication review)로 칭하기도 하는데, 우리가 흔히 말하는 사독이란 학술지 게재신청 논문에 대한 사전심사를 의미하기도 한다.²⁷⁾

사독은 문헌상의 기록으로 볼 때 그 역사가 300여년 가까이 거슬러 올라간다. 그러나 사독이 시작된 초기 시대에 있어서는 편집자와 심사위원의 책임이나 권한이 그리 크지 않았다. 기껏 투고된 논문 등의 주제가 학술지의 성격에 부합하는 지 여부만을 따졌을 뿐 연구논문의 완성도를 높이기 위한 노력은 거의 전적으로 저자의 몫이었다. 한동안 의학연

구 분야에서의 사독은 소극적 양상을 유지하여 왔다. 편집자와 심사위원의 권한이 강화된 오늘날의 사독제도가 틀을 갖춘 것은 2차 세계대전 이후부터인데 사독의 절차와 엄격성 등은 학술지마다 차이가 있었다.²⁸⁾ 의학연구 논문의 질 향상과 연구결과에 대한 사실상의 유일한 검증 수단으로 자리매김 된 사독제도의 효과와 신뢰성에 대한 논의는 1985년 Bailar와 Patterson²⁹⁾이 NEJM 독자논평란을 통해 문제를 제기한 것이 발단이 되었다. 이들이 제기한 문제의 중요성을 인식한 JAMA 편집진은 사독의 허와 실에 대한 과학적 평가를 위한 학술대회를 개최하기로 결정하였다. 그 결정에 따라 1989년 시카고에서 ‘제1차 International Congress on Peer review’가 열렸으며 대략 4년 주기로 계속되고 있다. 따라서 사독제도 자체의 효과와 효율성에 대한 논의가 시작된 1985년을 현대적 사독 제도 운영의 시점으로 간주하기도 한다.²⁸⁾

과거 그 효과에 대한 엄정한 검토도 없이 관행적으로 사독제도를 운영하였던 것은 그나마 안하는 것보다는 하는 것이 나을 것이라는 기대 때문이었다.²⁸⁾ 물론 1985년 이전에도 사독이 논문의 질 향상에 미치는 효과가 미미하고 비공정성 등 윤리적 측면에서의 문제점도 있다는 지적이 없었던 것은 아니다.³⁰⁾ 이러한 지적은 최근 시도되고 있는 사독의 효과평가 연구들에서 심도 있게 검토되고 있다.

Annals of Internal Medicine지에 투고된 논문의 사독 전후 질을 비교하였던 한 연구에서는 사독이 논문의 질 향상에 긍정적 효과가 있다고 주장하였다. 이 연구에서는 34가지 평가항목에 따라 사독 전후 논문의 질 향상 여부를 평가하였는데, 사독은 특히 연구의 제한점 서술, 연구결과의 일반화, 신중한 결론도출과 관련한 긍정적 효과가 있는 것으로 나타났다.³¹⁾ 그러나 Jefferson 등³²⁾의 문헌고찰 연구에서는 문장 표현의 교정 등 논문서술의 형식적 측면에서만 사독의 효과가 인정될 뿐 연구논문 자체의 본질에는 별다른 영향을 미치지 못한다고 하고 있다.

사독위원들마다 투고논문의 가치에 대한 판단이 다르다면 이는 분명 이상한 현상이다. 그러나 실제 논문사독 과정 중 흔하게 접하게 되는 문제이다. 이는 의학논문에서의 통계 활용과 관련한 사독을 할 때 더욱 극명해지는 경향이 있다. 두 개의 주요 신경과학 학술지 투고논문에 대한 사독위원들의 게재여부 판정 일치도를 평가한 연구결과에 따르면 일치도를 나타내는 kappa값이 0.08 정도인 것으로 나타났다. 이는 두 심사위원 간 판정의 일치수준이 우연에 의해 일치될 수 있는 수준에 불과하다는 것이며, 이렇게 판정 결과의 일치도가 낮은 주된 이유는 바로 통계활용의 적정성에 대한

평가의견의 차이 때문이었다.³³⁾

사독과정을 거치며 연구논문의 질을 향상시키기 위해서는 교정 가능한 연구의 결함을 찾아내어 저자에게 보완할 수 있는 기회를 주어야 한다. 이는 사독제도 운영의 가장 중요한 목적일 것이다. 이러한 사독 본연의 기능이 제대로 발휘되지 못하는 경우가 많기 때문에 사독의 효과에 대한 근원적인 의문이 제기되고 있는 것이다. 그러나 그 반대의 경우도 문제가 되기도 한다. 즉 연구의 결함이라고 보기 어려운 것 또는 연구의 제한점이기는 하지만 교정이 불가능한 사항을 집중 추궁하는 사독위원으로서의 바람직하지 못한 행태가 바로 그것이다. Bacchetti³⁴⁾는 사독위원들이 자주 지적하고 있는 “표본수 산정의 근거와 적정성”을 구체적인 사례로 꼽고 있다. 임상시험연구의 계획단계에서 대개는 적절한 표본수를 산정하게 된다. 특히 신약의 시판허가 등을 목적으로 한 임상시험연구라면 이 과정은 연구결과의 타당성을 높이기 위한 필수적인 절차로 중시되기도 한다. 그러나 적절한 표본수를 산정하기 위해서는 유의수준과 통계학적 검정력에 대한 합리적 가정을 해야 하고, 임상적으로 의미 있는 차이 또는 효과가 어느 정도인지에 대한 합의가 선행되어야 한다. 따라서 임상적으로 의미 있는 차이 등에 관한 신뢰할만한 정보를 얻을 수 없다면 표본수의 산정은 그야말로 숫자놀음에 불과한 것이다. 근거가 부족하여 표본수 산정을 사전에 못한 것은 연구의 제한점으로 받아들여야 하는 것이다. 무리하게 표본수의 적정성 문제를 지적하며 연구결과를 폄하하는 것은 사독의 취지에도 맞지 않는다. 사독위원의 무리한 지적으로 인해 연구자 스스로 연구의욕을 접거나 의미 없는 숫자놀음에 시간을 허비해서는 안 된다. 이와 유사한 또 다른 대표적 사례가 바로 최근 강조되고 있는 Intention-to-treat 원리의 미적용 문제이다. Intention-to-treat 원리의 요체는 연구 중도 탈락자들도 분석에서 배제하지 않아야 한다는 것이다. 그러나 핵심 결과변수에 대한 정보를 중도 탈락자들에게서도 정확히 얻어낼 수 없다면 이 원리의 적용은 곤란하다. 예를 들어 야뇨증 치료약물들의 효과를 비교하는 임상시험연구를 할 경우 약물의 부작용 또는 거주지역 변경 등으로 인해 중도 탈락하는 환자가 생길 수 있는데, 이들의 임상경과를 정확히 추적하는 것은 매우 어려운 일이다. 이런 경우 역으로 Intention-to-treat 원리를 적용하고자 한다면 그것 자체로 인하여 새로운 바이어스가 개입될 위험만 커진다.³⁵⁾

양질의 사독을 해줄 수 있는 외부 심사위원의 확보 여하에 따라 학술지 질 관리의 성패가 결정된다고 해도 과언은 아니다. Black 등³⁶⁾의 연구결과에 따르면 역학과 통계에 대한

교육을 받은 경험이 사독의 질과 밀접한 관련이 있다. 그러나 우수한 사독위원을 양성하거나 확보하는 것이 짧은 기간에 가능한 것도 아니고 상당한 자원을 투입하여야 하므로 학회 입장에서는 쉽지 않은 일이다. 그렇다고 해서 방법이 없는 것은 아니다. 외국의 몇몇 학회 또는 학술지에서는 잠재적인 사독위원들을 대상으로 자기주도학습 방식의 교육프로그램을 운영하고 있는데, 대표적인 사례가 BMJ에서 개발, 운영하고 있는 ‘Training package for BMJ peer reviewers (<http://resources.bmj.com/bmj/reviewers/training-materials>)’이다. BMJ의 잠재적인 사독위원들을 위해 개발된 이 온라인 프로그램은 누구나 별도의 회원가입 없이 무료로 이용할 수 있으며 사독에 대한 일반적 지침, 이론적 배경과 관련한 문헌, 사독 훈련용 논문 및 우수 사독 사례를 제공하고 있다.

통계학적 방법론 서술 및 결과 제시 원칙

국제의학학술지 편집인위원회의 ‘생의학 학술지에 투고하는 원고의 통일양식’에 포함되어 있는 통계 서술지침은 국내외 의학학술지 대부분에서 투고규정 또는 사독기준으로 채택되고 있다. 이 통일양식에서는 지난 1988년, 15개 항목으로 구성된 통계 서술지침을 명시하였던 것인데, 현재는 일부 내용이 삭제되어 대략 8개 항목 정도로 줄어든 상태이다.³⁷⁾ 이들 항목 하나하나의 의미를 살펴보고자 한다.

1. 투고 논문의 내용을 충분히 이해할 만한 독자가 연구결과를 재현할 수 있도록 통계학적 방법론을 자세하게 기술해야 한다.

의학학술지에 게재되는 논문이라면 그 학술지를 정기 구독하는 독자들이 한두 번 정도 정독함으로써 그 내용을 무난하게 이해할 수 있어야 한다. 또한 논문의 오류나 제한점에 대한 비판적 검토가 가능할 정도로 연구방법론이 충실하게 기술되어야 한다. 통계결과와 함께 통계학적 방법론에 대한 서술도 이러한 원칙을 따라야 한다는 것이다. 이를테면 정규성 여부 검토 등 주요변수에 대한 탐색과정, 연속변수의 범주화 방법, 대표값과 산포도 지표의 선택 기준, 무응답이나 열외값 처리 방법, 다변량분석에서의 독립변의 선정원칙과 모델의 적합도 평가방법 등을 충실하게 서술하여야 한다. 국내 의학학술지 게재논문들의 경우 대부분 사용한 통계프로그램과 적용한 통계기법을 단순 나열하는 수준인데, 이는 외국 의학학술지 게재논문들과는 다른 것이다. 통계학적 방법론의 서술은 꼭 ‘연구방법’란을 통해서만 할 필요는 없

다. 예를 들어 비모수검정기법인 Wilcoxon rank sums test 또는 Mann-Whitney U test를 이용하여 두 집단 간 평균값 차이에 대한 유의성 평가를 하였다면 왜 비모수검정 기법을 적용하였는지 그 근거를 제시하는 것이 바람직하다. 이럴 경우 “연구대상의 수가 적었고, 정규성 검정결과 정규 분포에 위배되는 것으로 확인됨에 따라 비모수검정인 Wilcoxon rank sums test 등을 시행하였다” 라고 ‘연구방법’란에 서술하거나, 표나 도표를 이용하여 그 결과를 제시하는 과정 중 유효표본수와 정규성 검정결과를 기재하는 것으로 대신할 수도 있다. 최근 다변량 로지스틱회귀분석이 질병발생 위험요인을 평가하거나 환자들의 예후인자를 규명하는 연구 등에 널리 활용되고 있다. 이런 분석기법을 제대로 적용하고자 한다면 독립변수 간의 관련성 검토, 합리적인 독립변수 선정, 모델의 적합도 등에 대한 엄정한 평가가 이루어져야 하며 그 내용이 서술되어야 한다. 결과변수와 의의학적 관련성이 인정되는 변수는 단변량 분석의 결과와 무관하게 다변량 로지스틱회귀분석의 독립변수로 포함할 수 있으나, 그렇지 않은 변수는 단변량 분석에서 P 값이 0.25 이하인 것을 포함하는 것이 원칙이다.³⁸⁾ 이런 원칙의 준수 여부를 서술하지 않거나, 단변량 분석에서 P 값이 0.05 이하인 것을 다변량 로지스틱회귀분석의 독립변수로 포함하였다고 서술하는 것은 타당하지 않다.

2. 가능하면 연구소견을 정량화하고 측정오차 또는 불확실의 정도(예를 들면 신뢰 구간 등)를 표시하는 지표를 같이 제시한다.

연구자료는 적절한 기술통계학적 분석과 통계검정 등 추측통계학적 분석 기법으로 처리되어야 하며, 그 결과는 가능한 정량적으로 제시되어야 한다. 또한 연구자료를 수집하기 위하여 사용된 수단 즉, 진단검사도구와 설문지 등의 정확도와 신뢰도에 대한 정보도 제시하여야 한다. 각 측정결과 및 측정도구의 정확도와 신뢰도에 대한 정보를 정확히 제시하는 것은 연구결과의 타당성 제고를 위해서도 필수적이지만 경우에 따라서는 후속연구자의 연구수행에도 직접적인 도움을 주기도 한다. 최근 들어 메타분석법이 임상의학분야 연구에 널리 활용되고 있다. 이 분석기법은 연구자마다 보고하는 결과가 상이하여 논란의 여지가 있거나, 전반적인 경향의 흐름은 인정됨에도 불구하고 개별연구에서는 표본수 등의 제한으로 인해 통계학적 유의성이 인정되지 않는 다수의 연구결과가 축적되고 있는 주제에 관한 계량적이고 통합적인 문헌고찰 수단이다.¹⁾ 따라서 연구결과를 정확하고 정량적으로 서술하는 것은 이 연구에 내재되어 있는 중요한 가치

를 후속연구자에 의해 발굴되도록 여지를 남기는 것과 다름이 없다. 통계검정의 결과도 $P < 0.05$ 또는 $P > 0.05$, “통계적으로 유의하였다” 또는 “통계적으로 유의하지는 않았다”와 같이 과거의 이분법적 판정 관례를 답습하지 말고 통계 프로그램에서 제시하고 있는 정확한 P 값(예를 들면 $P = 0.032$)을 제시하던가 아니면 95% 신뢰구간 등을 제시하여 독자 나름대로의 해석이 가능하도록 하여야 한다.

3. P 값만 제시하는 경우와 같이 통계검정 결과에만 의존하여 중요한 정량적 정보를 전달하지 못하는 일은 피해야 한다.

거의 대부분 표본조사 방식으로 진행되는 의학연구의 특성 상 연구결과를 일반화하기 위해서는 통계검정 등 추측통계학적 분석을 하여야 한다. 이러한 추측통계학적 분석의 목적은 단순히 비교하고자 하는 집단 간 차이의 유의성을 이분법적으로 판정하는 것만으로 달성되지는 않는다. 표본의 결과를 일반화하고자 하는 모집단의 특성을 가급적 구체적으로 도출해 내는 것이 필요하다. 그렇기 때문에 “유의성이 있다 혹은 없다”와 같은 통계검정 분석기법의 이분법적 결론보다는 비교집단 간 차이의 정도와 우연에 의한 변이의 수준을 양적으로 표현하는 95% 신뢰구간이 최근 선호되고 있는 것이다.

4. 연구설계나 통계학적 방법론에 관한 참고문헌을 제시하는 것이 좋다.

국내 의학학술지 지면상에서의 통계요류는 외국 의학학술지의 경우보다 흔한 것은 분명하다. 통계학적 방법론을 의학연구의 핵심적인 방법론으로서 중시하는 것이 아니라 번거롭고 지루한 수학적인 계산과정으로 간주하는 경향이 큰 우리나라 의학연구자들의 안이한 인식의 산물이기도 하다. 한편, 대다수 의학연구자들이 논문작성 과정 중 가장 큰 어려움을 겪는 단계가 바로 통계분석 과정이다. 이런 어려움을 해결하기 위해서는 통계 전문서적을 찾아 볼 수도 있겠지만, 유사한 선행연구에서의 통계학적 방법론 적용사례를 검토하는 것이 더욱 효율적일 수 있다. 그렇다면 통계학적 방법론의 활용을 위하여 참고한 선행연구를 참고문헌으로 제시하는 것은 당연하다. 그러나 국내 의학학술지 게재논문들에서는 통계학적 방법론을 참고한 문헌을 참고문헌으로 제시하는 경우는 별로 없다. 외국 의학학술지 게재논문들의 경우는 다르다. 웬만한 통계지식을 갖고 있는 의학연구자라도 연구논문의 작성을 위한 자료를 분석할 때 통계서적과 함께 선행연구에서의 분석방법을 참고하게 된다. 따라서 통계학

적 방법론을 활용한 연구논문을 발표할 때에는 연구방법론의 과학적 근거를 제시한다는 차원에서 반드시 통계학적 방법론에 대한 참고문헌을 충실하게 기재하여야 한다. 때로는 연구논문에서 제시하고 있는 의학적 결론 그 자체보다도 연구방법론에 대한 참고문헌이 독자들과 후속연구자에게 더욱 큰 관심의 대상이 될 수 있다는 점에서 통계학적 방법론과 관련한 참고문헌도 가능한 구체적으로 기재하여야 한다.

5. 통계 용어, 약어, 기호의 의미를 명확히 정의하며 활용하여야 한다.

의학연구자라면 누구나 혼동하지 않고 알고 있을 만한 통계 용어, 약어, 기호가 아니라면 그 정의를 명확히 제시하며 활용하여야 한다는 것이다. 예를 들어 소아알레르기 호흡기학회지 게재논문들에서 드물지 않게 언급되고 있는 민감도와 특이도 등의 경우 대다수 독자들이 그 개략적인 의미만 알지 산출공식과 정확한 개념까지 이해하지 못하는 경우가 많다. 이들 용어와 같이 사용되기도 하는 양성예측도와 음성예측도를 의미하는 PV(+)나 PV(-) 등의 약어는 공인된 통일용어라고 할 수도 없다. 이런 용어들의 산출공식과 의미가 명확히 정의되며 사용되어야 한다는 것이다.

6. 어떤 컴퓨터 통계프로그램을 사용하였는지를 명시하여야 한다.

지금처럼 의학연구의 규모가 대형화되고 속도가 빨라지게 된 것은 컴퓨터와 통계프로그램의 발전 덕분에 가능해진 것이다. Statistical Analysis System (SAS)와 Statistical Packages for Social Science (SPSS) 등과 같은 대표적인 통계프로그램은 이제 의학연구자들에게 없어서는 안 될 도구가 되어 버렸다. 이와 같은 통계프로그램은 그 편리성에 대해서는 의심의 여지가 없으나, 유용성에 관해서는 일부 비판도 제기되고 있다. 통계프로그램은 번잡한 수학적 계산으로부터 많은 연구자들을 자유롭게 하였으나 연구문제에 대한 고민이나 자료에 대한 탐색마저도 소홀히 하게 되는 역기능도 있기 때문이다. 또한 통계프로그램이라는 것도 일부 결함이 있을 수 있어 동일한 자료에 대한 분석결과가 프로그램에 따라 다소간 차이를 보일 수도 있으며, 일부 분석기법에 대해서는 특정 프로그램이 훨씬 편리하고 강력한 기능을 갖기도 한다. 예를 들어 진단검사도구의 정확도를 평가하는데 널리 활용되는 receiver operating characteristic (ROC) 곡선 분석은 Medcalc 프로그램을 통해서 하는 것이 가장 편리하다는 것이 대표적인 예이다. 그러므로 통계프로그램을 활용한 연구자는 자신이 이용한 통계프로그램의 종

류, 버전(version), 구체적인 통계기법 적용을 하는데 참고한 매뉴얼 등을 제시하여야 하며, 만약 통계프로그램을 선택하는데 특별한 이유가 있었다면 그 이유도 설명하여 후속연구자들에게 참고가 될 수 있도록 배려하는 것이 바람직하다.

7. ‘연구결과’란에 통계결과를 서술할 때 그 결과를 도출한 통계기법을 명시하여야 하며, 표나 그림에 기술한 값 모두를 본문에서 언급할 필요는 없다.

만약 해당 연구논문에서 카이제곱검정이라는 한 가지 통계기법만을 사용하였고 이를 ‘연구방법’란에 서술하였다면, ‘연구결과’ 중 통계검정 결과를 제시하며 굳이 이것이 카이제곱검정에 의한 것인지를 명시할 필요는 없을 수도 있다. 그러나 핵심적인 연구결과와 관련한 정보를 구체적으로 기술하고자 한다면 χ^2 값이나 자유도를 제시하고 P 값이 카이제곱검정에 의하여 산출되었다고 명시하는 것이 더욱 바람직하다. 반면, 한 논문에서 여러 가지 통계기법을 적용한 경우라면 ‘연구결과’란에서 제시하고 있는 다양한 통계분석 결과 하나 하나가 구체적으로 어떤 통계기법에 의한 결과인지를 기재하여야 한다. 그러나 표나 그림에 기술한 값 거의 대부분을 본문에 기술하거나, 표에 포함된 결과를 그림으로 중복 제시하는 것은 지면의 낭비이다.

8. 무작위(random), 정상(normal), 유의한(significant), 상관성(correlation), 표본(sample)과 같은 전문용어를 임의로 사용해서는 안된다.

정확하고 적절한 통계용어의 사용은 연구논문의 과학성 제고라는 측면에서 강조되어야 한다. 국내 의학연구자들이 흔히 사용하기도 하는 정상 범위(normal range)의 경우가 전문용어를 임의로 사용하는 대표적인 사례이기도 하다. 통계학적으로 ‘정상(normal)’이란 변수 분포의 정규성을 의미하며, 범위(range)의 경우도 특정 산포도 지표를 칭하는 것이다. 따라서 정상범위보다는 참조구간(reference interval)이라는 용어의 사용이 권장되고 있다.³⁹⁾ 소아알레르기 호흡기학회지만 하더라도 동일한 통계용어가 논문마다 달리 사용되는 경우를 드물지 않게 볼 수 있다.

소아알레르기 호흡기학회지 게재논문들에서의 통계활용 현황

소아알레르기 호흡기학회지 게재논문들의 통계학적 방법론 활용과 서술 현황을 평가해보았다. 2008-2010년 중 발간된 제18권-제20권에 게재되었던 원저 형식의 논문 101편을 대상으로 하였다. 이들 논문들에서의 통계활용의 범위, 사용된 통계기법, 그리고 국제의학학술지 편집인위원회 '생의학 학술지에 투고하는 원고의 통일양식'에 포함되어 있는 통계 서술지침과의 부합성을 검토한 것이다.

우선 이들 논문들에서의 통계활용 범위와 개별 기법의 활용빈도를 요약하면 Table 1에서 보는 바와 같다. 통계분석은 크게 기술통계학적 분석과 통계검정 등 추측통계학적 분석으로 구분되는데, 추측통계학적 분석을 하기 위해서는 기술통계학적 분석 과정을 거쳐야 한다. 따라서 기술통계학적 분석을 했는지 여하에 따라 통계분석 여부를 판단할 수 있는 것이다. 검토대상 논문 101편 중 84.2%인 85편에서 기술통계학적 분석을 하여 그 결과를 제시하고 있었다. 이들

85편의 논문들에서는 평균(78.8%), 표준편차(69.4%), 빈도 또는 백분율(60.0%) 등을 대표값과 산포도 지표로서 흔히 활용하고 있었다. 추측통계학적 분석을 시행하였거나 진단검사도구의 정확도 평가 기법 등을 활용한 경우는 전체의 83.5%인 73편이었다. 이 논문들에서 가장 빈번하게 활용된 통계기법은 카이제곱검정(59.2%)이었다. 그 다음으로 Student's *t*-test(43.7%), 비모수검정(28.2%), Pearson 상관분석(21.1%) 순이었다. 로지스틱회귀분석은 9편(12.7%)의 논문에서만 활용되고 있었고, ROC 곡선 분석 등을 통해 진단검사도구의 정확도를 평가한 논문은 4편(1.4%)이 있었다. 소아알레르기 호흡기학회지 게재논문들에서는 기초 통계기법들이 주로 활용되고 있음을 보여주는 결과이다.

'생의학 학술지에 투고하는 원고의 통일양식'에 규정되어 있는 통계 서술지침의 준수상태를 검토해 본 결과 몇 가지 개선되어야 할 문제점이 나타났다. 통계기법을 활용한 85편의 논문 중 '연구방법'란 등에 통계학적 방법론을 서술하지 않은 경우가 19편(22.4%) 있었으며, 통계학적 방법론을 서술한 66편의 논문 중 6편(9.1%)은 서술내용이 제시된 결과와 일치하지 않고 있었다. 추측통계학적 분석기법으로 가장 많이 활용된 카이제곱검정은 교차표상 전체 칸의 20% 이상에서 기대도수가 5 이하라면 그대로 적용해서는 안된다.¹⁾ 그러나 카이제곱검정을 활용한 논문 42편 중 10편(23.8%)에서 기대도수가 5 이하인 칸이 20% 이상이어서 부적절한 통계기법 적용의 오류가 많을 것으로 판단되고 있다. 추측통계학적 분석기법을 활용한 71편의 논문 중 신뢰구간을 제시한 논문은 15편(21.1%)에 불과하였으며, 나머지 61편은 이분법적 통계검정 결과에 의존하여 연구결론을 도출하고 있었다. 통계분석 결과를 제시하면서 각 결과의 유효 표본수를 명확히 제시한 논문은 26편(36.6%)이었으며, 각 통계검정 결과가 구체적으로 어떤 분석기법에 의하여 산출된 것인지를 빠짐없이 기술한 논문은 2편(2.8%) 뿐이었다.

결 론

이상과 같이 의학학술지 게재논문들에서의 통계활용 및 통계학적 오류 실태, 의사들의 통계학적 방법론 이해도 평가 결과, 의학학술지 게재논문들의 질 향상을 위한 통계사독의 현황, 의학학술지 투고논문들이 준수하여야 할 통계학적 방법론 서술 및 결과제시 원칙, 소아알레르기 호흡기학회지 게재논문들에서의 통계활용 현황을 검토해보았다. 이 검토결

Table 1. Utilization of Statistical Methods in 101 Original Articles in Pediatric Allergy and Respiratory Disease (2008-2010)

Type of statistical method	Value
Descriptive statistics	85 (84.2)
Mean	67 (78.8)
Median	4 (4.7)
Standard deviation	59 (69.4)
Range	10 (11.8)
Frequency or percent	51 (60.0)
Inferential statistics + other statistics	71 (83.5)
Student's <i>t</i> -test	31 (43.7)
ANOVA with multiple comparisons	12 (16.9)
Paired <i>t</i> -test	4 (5.6)
Non-parametric test	20 (28.2)
Chi-square test	42 (59.2)
Fisher's exact test	11 (15.5)
Pearson correlation analysis	15 (21.1)
Spearman correlation analysis	7 (9.9)
Linear regression analysis	3 (4.2)
Logistic regression analysis	9 (12.7)
Survival analysis	1 (1.4)
Sensitivity/Specificity, ROC curve analysis	4 (5.6)

Values are presented as number (%). ANOVA, analysis of variance; ROC, receiver operating characteristic.

과를 토대로 소아알레르기 호흡기학회지의 질적 발전을 위한 몇 가지 제언을 하는 것으로 결론을 맺고자 한다.

1. 통계서술과 관련한 투고규정의 개정

구체적이고 명확한 논문작성 지침을 제시하는 것은 그 자체만으로도 연구논문의 질 향상 효과가 있다. 또한 사독위원의 노력과 시간도 경감되는 부수적 효과도 기대할 수 있고 학회 소속 연구자의 국제 경쟁력을 높이는 데도 도움이 된다. 심사해야 할 투고논문이 매우 많은 외국 주요 의학학술지의 경우 편집진이 논문의 초록만을 검토한 후 본격적인 사독여부를 판단하기도 한다. 이에 따라 국제의학학술지 편집인위원회에서도 투고논문 초록의 완결성을 강조하고 있는 것이다.³⁸⁾ 그러나 국내 의학학술지에서는 초록의 중요성을 간과하는 경향이 있다. 소아알레르기 호흡기학회지 투고규정만 보더라도 초록의 분량과 형식 정도만을 간략하게 언급하고 있을 뿐, 핵심적인 연구결과의 정확한 기재와 본문내용과의 일치 여부 확인 등을 저자에게 요구하지는 않고 있다. 또한 소아알레르기 호흡기학회지 투고규정 상 통계에 관한 서술지침은 “통계처리법에 관한 설명 시, 사용된 통계방법에 대한 정확한 기재와 논문에서 사용된 의미 있는 P 값의 구간을 설명한다.(예: $P<0.05$) 그러나 최근 P 값에 관해 계산된 P 값을 직접 표시하는 경향이 있어 이를 사용해도 무방하다.(예: $P=0.004$) P 자는 대문자 이탤릭체로 표기한다”가 전부이다. 국제의학학술지 편집인위원회의 ‘생의학 학술지에 투고하는 원고의 통일양식’에 포함되어 있는 통계 서술지침에서 강조하고 있는 중요내용의 상당부분이 누락된 것이며, 특히 P 값의 경우 산출한 값을 그대로 기재하는 것이 최근의 추세임을 고려할 때 지침으로서의 타당성이 낮다고 할 수 있다.

2. 학회 회원 등을 대상으로 한 통계학적 방법론 교육 활성화

최근 소아알레르기 호흡기학회지에 게재된 논문의 상당수에서 통계학적 오류가 있거나 보편적인 통계 서술지침을 위반하고 있는 것으로 확인되고 있다. 연구자 대상의 통계학적 방법론 교육프로그램 운영이 필요함을 시사하는 결과이다. 학회기간 중 실용적인 내용의 통계강좌를 운영하거나, 학술지 지면을 활용하여 정기적으로 지면강좌를 실시하는 것은 학회의 발전과 미래의 우수한 사독위원을 확보하기 위한 효율적인 투자가 될 것이다. 젊은 연구자들의 연구방법론에 대한 동기부여를 위하여 전공의 학습목표와 전문의 시험 문항에 통계학적 방법론을 포함한 연구방법론 관련 내용을

포함하는 조치도 추진해볼 만하다.

3. 현행 사독제도의 보완과 사독위원회에 대한 보상 강화

외국 의학학술지에서 시도되고 있는 사독의 질에 대한 지속적인 모니터링이 필요하다. 사독위원이 자의적 판단기준에 따라 사독을 하게 된다면 논문의 질 향상이라는 본연의 목적달성도 어렵고 연구자들의 후속연구 의욕마저도 꺾게 된다. 따라서 사독위원회에 대한 교육과 활동지원이 뒷받침되어야 한다. 사독 지침서와 모범사독 사례집을 발간하게 된다면 긍정적 과급효과가 클 것이다. 또한 사독위원들에 대한 동기부여를 하고, 노고에 상응하는 보상을 제공하는 것은 사독의 내실화를 촉진하기 위한 전제가 됨을 인식하여야 한다. 현재와 같이 사독위원의 봉사와 희생에 전적으로 의존하는 것은 곤란하다. 여기서 말하는 보상이란 사독위원회에 대한 금전적 인센티브만을 의미하는 것은 아니다. 사독이 학회 회원으로서 당연히 하여야 하는 봉사활동의 수준을 넘어 그 자체가 의미 있는 학술활동으로 인정되어야 한다. 연구실적, 연구비 신청자격, 승진심사를 위한 업적으로 반영되는 것이 당연하다.⁴⁰⁾ 한편, 사독위원으로서 의학발전에 헌신하고 계신 분들에 대한 예우 차원에서라도 사독 학술상을 시상하는 방안을 검토해볼 만하다. 이 상은 사독의 질이 가장 우수한 분과 사독건수가 가장 많은 분을 대상으로 수여할 수 있다. 그 중 사독의 질은 논문게재가 확정된 저자들로 하여금 평가토록 하는 것이 바람직하다. 사독의 실질적인 질 향상은 물론 사독과정 중 연구자와 사독위원들이 대등한 위치에서 건설적인 의견교환을 가능하게 한다는 장점이 있기 때문이다.

참 고 문 헌

1. Altman DG. Practical statistics for medical research. 2nd ed. Boca Raton: Chapman & Hall/CRC, 1999.
2. 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.
3. Schor S, Karten I. Statistical evaluation of medical journal manuscripts. JAMA 1996;195:1123-8.
4. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics

- and results in the medical literature. *JAMA* 2007;298:1010-22.
5. Polychronopoulou A, Eliades T, Taoufik K, Papadopoulos MA, Athanasiou AE. Knowledge of European orthodontic postgraduate students on biostatistics. *Eur J Orthod* 2011;33:434-40.
6. Windish DM. Brief curriculum to teach residents study design and biostatistics. *Evid Based Med* 2011;16:100-4.
7. Koh WJ, Lee SJ, Kang MJ, Lee HJ. An evaluation of the use of statistical methods in the *Journal of Tuberculosis and Respiratory Diseases*. *Tuberc Respir Dis* 2004;57:168-79.
8. 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.
9. White SJ. Statistical errors in papers in the *British Journal of Psychiatry*. *Br J Psychiatry* 1979;135:336-42.
10. Felson DT, Cupples LA, Meenan RF. Misuse of statistical methods in *Arthritis and Rheumatism*. 1982 versus 1967-68. *Arthritis Rheum* 1984;27:1018-22.
11. Bailar JC 3rd, Mosteller F. Guidelines for statistical reporting in articles for medical journals. amplifications and explanations. *Ann Intern Med* 1988;108:266-73.
12. Stell IM, Gransden WR. Simple tests for septic bursitis: comparative study. *BMJ* 1998;316:1877.
13. Deeks JJ, Altman DG. Sensitivity and specificity and their confidence intervals cannot exceed 100%. *BMJ* 1999;318:193-4.
14. Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. *JAMA* 1999;281:1110-1.
15. Khan KS, Chien PF, Dwarakanath LS. Logistic regression models in obstetrics and gynecology literature. *Obstet Gynecol* 1999;93:1014-20.
16. Kuo YH. Extrapolation of correlation between 2 variables in 4 general medical journals. *JAMA* 2002;287:2815-7.
17. Wang Q, Zhang B. Research design and statistical methods in Chinese medical journals. *JAMA* 1998;280:283-5.
18. Olsen CH. Review of the use of statistics in infection and immunity. *Infect Immun* 2003;71:6689-92.
19. 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.
20. Son YI, Cho YS, Kim HY, Ha BS, Kim SI, Shin MH. The use of statistical analysis: in the Korean Journal of Otolaryngology-Head and Neck Surgery. *Korean J Otorhinolaryngol-Head Neck Surg* 1999;42:1364-9.
21. Hwang K, Lee HJ, Kim YJ, Lee SI. Statistical errors in papers in the *Journal of Korean Society of Plastic and Reconstructive Surgeons (JKPRS)*. *J Korean Soc Plast Reconstr Surg* 2001;28:302-9.
22. Friedman SB, Phillips S. What's the difference? Pediatric residents and their inaccurate concepts regarding statistics. *Pediatrics* 1981;68:644-6.
23. Berwick DM, Fineberg HV, Weinstein MC. When doctors meet numbers. *Am J Med* 1981;71:991-8.
24. Reznick RK, Dawson-Saunders E, Folse JR. A rationale for the teaching of statistics to surgical residents. *Surgery* 1987;101:611-7.
25. Cheatham ML. A structured curriculum for improved resident education in statistics. *Am Surg* 2000;66:585-8.
26. Ambrosius WT, Manatunga AK. Intensive short courses in biostatistics for fellows and physicians. *Stat Med* 2002;21:2739-56.
27. Wager E, Godlee F, Jefferson T. How to survive peer review [Internet]. London: BMJ Books, 2002 [cited 2011 Sep 9]. Available from: <http://resources.bmj.com/bmj/pdfs/wager.pdf>.
28. Rennie D. Editorial peer review: its development and rationale. In: Godlee F, Jefferson T, editors. *Peer review in health sciences* [Internet]. London: BMJ Books, 1999 [cited 2011 Sep 9]. Available from: <http://resources.bmj.com/bmj/pdfs/rennie.pdf>.
29. Bailar JC 3rd, Patterson K. The need for a research agenda. *N Engl J Med* 1985;312:654-7.
30. Goldbeck-Wood S. Evidence on peer review—scientific quality control or smokescreen? *BMJ* 1999;318:44-5.
31. Goodman SN, Berlin J, Fletcher SW, Fletcher RH. Manuscript quality before and after peer

- review and editing at *Annals of Internal Medicine*. *Ann Intern Med* 1994;121:11-21.
32. Jefferson T, Alderson P, Wager E, Davidoff F. Effects of editorial peer review: a systematic review. *JAMA* 2002;287:2784-6.
33. Rothwell PM, Martyn CN. Reproducibility of peer review in clinical neuroscience. is agreement between reviewers any greater than would be expected by chance alone? *Brain* 2000;123:1964-9.
34. Bacchetti P. Peer review of statistics in medical research: the other problem. *BMJ* 2002; 324:1271-3.
35. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ* 2001;165:1339-41.
36. Black N, van Rooyen S, Godlee F, Smith R, Evans S. What makes a good reviewer and a good review for a general medical journal? *JAMA* 1998;280:231-3.
37. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publications [Internet]. [place unknown]: ICMJE; c2009 [cited Sep 9]. Available from: <http://www.icmje.org/>.
38. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. New York: Wiley, 2000.
39. Bland M. *An introduction to medical statistics*. 3rd ed. Oxford: Oxford University Press, 2000: 279.
40. Davidoff F. Improving peer review: who's responsible? *BMJ* 2004;328:657-8.