

대한치과의사협회지에 게재된 연구논문의 과학적 방법과 임상적 적용성에 대한 분석

미국 메이요클리닉 보철과
최 용 근

ABSTRACT

Systematic analysis of the scientific methodology and clinical relevance of original articles published in the Journal of Korean Dental Association

Mayo Clinic Rochester, Department of Dental Specialties, Section of Prosthodontics.
Yong-Geun Choi, DDS, MPH, MPH

Voices from Korean dentists expressing their limited understanding and lack of clinical pertinence of the original articles in Journal of Korean Dental Association exist. This study intended to find the possible cause(s) of the dissatisfaction of Korean private dentists and to suggest the future direction for the better use of the research articles in Journal of Korean Dental Association as a major practical source of clinical knowledge. 36 original articles published during 2 years (January 1999 through December 2000) in the Korean Dental Association website were analyzed using clinical epidemiological methods. The results show that most original articles have imperfections in scientific methodology as well as statistical analysis. In addition, clinical pertinence was not clearly expressed in most original articles. This study concludes that Korean dentists' limited understanding and difficulties in applying the articles' results to the real clinical problems are mainly due to the articles' deficiency of scientific methodology, improper statistical analysis and lack of clinical pertinence. This study, therefore, suggests that future original articles be made in compliance with scientific research methodology as well as principles of statistical analysis. The continuing education regarding critical appraisal skills in reading original articles should be also provided to Korean private dentists.

Key Words : clinical epidemiology, clinical pertinence, scientific methodology, statistical analysis, original articles in the Journal of Korean Dental Association

서론

한국에 있는 대부분의 대한민국의 치과의사가 받아보는 대한치과의사협회지는, 특히 개원치과의사들에게는 학술적인 정보를 접할수 있는 가장 보편적인 매체일 것이다. 따라서 매월 발행되는 대한치과의사협회지는 매년 10시간이내의 보수교육보다도 훨씬 더 효과적인 교육적(학습적)기능을 담당할 수 있어야 하는데도 불구하고, 대한치과의사 협회지의 독자의 소리난에는 현재 개원치과의사들이 대한치과의사협회지에 게재된 연구논문이 난해하거나 임상에 적용하기가 어렵다는 점을 표현하고 있다⁽¹⁾.

따라서 이 논문의 목적은 대한치과의사협회지에 게재된 연구논문들을 읽으면서 개원치과의사들이 난해하게 느끼거나 임상에 적용하는데 어려움을 느끼는 원인이 어디에 있는가 하는 것을 밝혀냄으로써, 앞으로 대한치과의사협회지에 게재되는 연구논문의 내용을 개원치과의사들이 임상에 효과적으로 활용할 수 있는 방안을 제시하는 데 있다.

연구 재료 및 방법

1999년 1월부터 2000년 12월까지 대한치과의사협회지에 게재되어 대한치과의사협회 인터넷 홈페이지에 올려져 있는 연구논문(원저 또는 Original Article) 총 41편중에서 데이터가 없는 논문 2편과 임상증례(case report)에 해당하는 논문 1편, 그리고 논문의 요약(abstract)만 있는 논문 1편은 제외하고 나머지 36편의 연구논문을 대상으로 하여, 1) 연구방법, 데이터의 통계분석방법, 고찰 부분이 과학적 연구방법과 통계학적 분석 및 추론의 원칙들에 얼마나 합당한지를 분석하였고, 2) 연구논문들의 임상적 적용성을 측정하기 위하여 연구목적과 연구결론부분의 내용이 임상에 적용할 수 있는 가능성에 얼마나 접근하려 하였는가를 평가하였으며, 3) 연구재료의 종류도 분석하였다.

대한치과의사협회지에 1954년부터 2000년 12월까지 1163편의 연구논문(Original Article)이 게재되었

는데⁽²⁾, 본 연구에서는 그중 2년동안 게재된 36편만을 표본으로 선정하였으므로, 이 3.2% (36/1163)의 표본이 제공하는 정보를 근거로 하여, 위에서 제시한 평가기준들에 합당한 논문들이 차지하는 비율(Proportion)이 전체 1163편중에서는 얼마나 될까 하는 것을 추정(estimation)하기 위하여, 모집단비율(Population proportion)의 95%신뢰구간을 계산하였다. 계산방법은 통계프로그램을 사용하지 않고 공식을 이용하여 전자계산기로 직접 계산하였다. 또한, 연구재료의 종류를 조사한 항목에서는 연구 재료의 종류별로 모집단(1163편의 논문)에서 차이가 있는가를 검증하기 위하여 Chi-square test를 사용하였고 계산방법은 역시 수작업으로 하였다⁽³⁾. 이 연구에서 사용한 표본수가 30편 이상이므로 추정의 전제조건(assumption)은 충분히 만족시키는 것을 계산 전에 미리 확인하였다.

연구 결과

표 1. 논문의 수 및 연구주제분야

논문의 주제 분야 기초 : 임상	비율 5 : 31 (14% : 86%)	95% 신뢰구간 [0.75 - 0.97]
----------------------	-------------------------------	---------------------------

표1. 논문의 수 및 연구주제분야 표1 에 의하면, 게재된 연구논문의 대부분이 임상분야를 다루고 있었고 총 연구논문 1163편에서도 최하 75% 최고 97%가 임상분야의 연구논문임을 추정할 수 있다. 따라서 대한치과의사협회지의 대다수 독자층은 임상을 주로 하는 개원치과의사라는 것을 고려할 때 게재된 연구논문의 소재가 대부분 임상분야라는 점은 적절하다고 볼 수 있다.

표2 는 31편의 임상연구논문이 주로 구강외과, 교정, 그리고 임플란트에 관한 연구로 편중 되어 있는 것을 보여주고 있다. 이 데이터는 대부분의 개원치과의사는 치과의 모든영역의 환자를 돌보는 것을 고려할 때, 앞으로 게재된 논문의 주제분야는 각 임상영역별로 균형을 이루도록 조정이 필요하다는 것을 보여주고 있다.

표2. 임상연구분야별로 구분한 논문의 빈도수

구강외과	치주	보건의료	보철	코골이
8	2	2	1	1
교정	악관절	근관치료	방사선	합계
6	2	1	1	31
임플란트	병리	골이식	소아치과	
3	2	1	1	

표3의 윗칸의 데이터에 의하면, 연구재료는 주로 인체를 다루고 있으며, 각 종류별로 관찰된 비율의 차이가 우연에 의해 일어날 확률도 0.5% 이하이기 때문에 각 종류간에는 유의한 차이가 있다고 할 수 있다. 따라서 대한치과의사협회지에는 연구결과를 인체에 적용 또는 일반화(extrapolation)시킬 수 있는 직접적인 임상적 가치가 있는 연구논문이 주로 실린다고 할 수 있다. 하지만, 표3의 아래칸의 데이터는 연구재료의 선정과정 및 특성이 명확하지 않은 논문이 대부분을 차지하는 것을 보여주고 있다. 전체 1163편의 논문에서는 적어도 68% 많게는 94%의 논문이 연구재료의 선정과정이나 특성이 불분명하다는 것을 추정할 수 있다.

표3. 연구재료의 종류와 선정과정 및 특성

연구 재료의 종류 (인체 : 동물 : 무생물)	비율 29 : 5 : 2 (80% : 14% : 6%)	Chi-square = 36.5 (degree of freedom : 2) p-value < 0.005
연구재료의 선정과정 및 특성이 명확히 제 시되지 않은 논문 : 연 구재료의 선정과정 및 특성이 명확히 제시된 논문	비율 29 : 7 (81% : 19%)	95% 신뢰구간 [0.68 - 0.94]

따라서 표3을 종합적으로 풀이하면, 연구논문의 재료는 주로 인체이지만 표본의 선정과정 및 표본의 특성이 불분명하기 때문에, 독자들이 연구논문을 읽고 나서 그 연구결과를 여러계층의 인간집단(성별, 지역별, 사회계층별)중에서 특히 어떤특성을 가지는 인간집단에게 적용해야하는지 알 수 없는 연구논문이 대다수라는 것을 알 수 있다.

표4는, 데이터를 계측하는 과정에서 측정자간에 계측의 일관성(consistency)이 있다는 것과, 측정자가

표4. 연구방법

연구데이터의 계측과정에서 일관성이 있으며 계통적 오류(bias)는 없다는 것이 입증되지 않은 논문 : 연구데이터의 계측과정에서 일관성이 있으며 계통적오류(bias)는 없다는 것이 입증된 논문	비율 28 : 8 (78% : 22%)	95% 신뢰구간 [0.64 - 0.92]
--	-----------------------------	---------------------------

계측하는 데이터나 연구대상자가 제공하는 데이터에 계통오차(bias)가 없다는 것이 입증되지 않은 논문이 대다수이고, 따라서 전체 1163편의 논문에서도 적게는 64% 많게는 92%까지가 이런 논문에 해당된다고 추정할 수 있다. 데이터의 산출과정이 일관성이 없는 것은 연구결과가 통계적으로 유의성이 없다는 방향으로 잘못 나타나는 경향이 있고, 계통오차(bias)가 있는 논문은 실제로는 유의성이 없는 연구결과를 유의성이 있는 의미있는 연구결과로 변질시키는 부작용이 있으므로, 표4에 의하면 연구논문의 상당수가 연구결과를 신뢰하기 어려운 상태라는 것을 알 수 있다. 특히 잘못된 연구결과는 임상가들에게 잘못된 지식을 제공하여 그 피해는 결국 환자에게 간다는 점을 고려할 때, 앞으로 게재되는 논문은 데이터 계측의 일관성 및 정확성을 확보해야 할 필요성이 있다.

표5. 통계분석방법

통계분석방법의 종류	빈도수 : proportion : 11 t-test : 4 paired t-test : 5 ANOVA : 7 chi-square test : 5 Kruskal-Wallis : 1 판별분석 : 1
통계분석결과에 대한 해석이 적절하지 않은 논문 : 통계분석결과에 대한 해석이 적절한 논문	비율 32 : 4 (89% : 11%) 95% 신뢰구간 [0.79 - 0.99]

표5는 여러 가지 통계분석방법중에서 서술적통계에 해당하는 비율(proportion)이 많이 사용되고 있다는 것을 보여준다. 이표는 연구논문의 상당수가 분석연구(analytical study)보다는 관찰된 데이터를 분석

없이 그대로 보고하는 기술연구(descriptive study)라는 것을 나타내주는 자료로서, 표본추출과정에서 데이터에 개입되는 무작위오차(random error)를 분리하기 위하여⁽⁴⁾ 앞으로 게재되는 연구논문의 데이터는 통계적분석을 반드시 거쳐야 될 필요성을 의미한다.

또한, 이 표에는 나타나 있지 않지만 연구방법이나 변수의 특성을 고려할 때 12편의 논문에서 통계분석 방법의 선택이 부적절하였고, 주로 사용하지 않는 희귀한 통계분석방법을 채택하였기 때문에 통계학에 대한 전문적인 교육을 받지 않은 개원임상가들이 논문을 이해하는데 어려움이 있을 가능성이 높은 논문도 1편 있었다.

표5는, 통계분석중에서 가장 심각한 문제는, 통계분석결과에 대한 해석이 정확하지 않은 논문이 89%나 되며, 전체 1163편의 논문중에서는 적어도 79% 많으면 전부(99%)가 통계분석의 결과를 잘못 해석하고 있는 것을 보여주고 있다. 그 유형별로는 통계적 유의성에 대한 해석을 반대로 하는 경우(p-value에 대한 해석을 거꾸로 해서 유의한 차이가 있는데 없다고 하고, 없는데 있다고 하는 경우), 유의성이 없을 때 표본의 크기를 고려하지 않은 해석, 통계적용어를 잘못 사용하는 경우 (예 : 95% 유의수준), p-value가 0.05보다 커서 통계적으로 유의하지 않을 때 제2종 오류(Type-II error)에 대한 정보없이 귀무가설(null hypothesis)을 받아들이는 경우가 있었다. 또한 빈번한 예로서, p-value는 표본의 크기에 큰 영향을 받으므로 이것을 평가할 수 있는 신뢰구간을 제시하는 것이 필요한데도 불구하고, 대다수의 논문에는 신뢰구간(confidence interval)이 보고되지 않았다⁽⁵⁾.

표6. 총괄 및 고찰의 내용

타 연구의 고찰에 치중한 논문 수 : 자기 연구의 방법에 대한 분석과 평가	비율 32 : 4 (89% : 11%)	95% 신뢰구간 [0.79 - 0.99]
---	-----------------------------	---------------------------

논문의 연구결과부분에서 보고된 데이터가 과학적으로 신뢰할 만하다는 것을 입증하기 위해서는, 표본추출과정 및 데이터의 측정과정에서 변이

(variability) 및 계통오차(bias)가 개입되지 않았거나 또는 그 개입정도가 데이터에 미치는 영향이 매우 적다는 것을 분석하고 평가하여야 하는 데⁽⁶⁾, 표6에 의하면, 게재된 대부분의 논문들의 총괄 및 고찰부분의 주된 내용은 기존에 나와 있는 다른 연구자들의 연구결과와 일치하는 것을 언급함으로써 연구결과부분에서 보고된 데이터의 신뢰성을 입증하려는 것을 보여준다. 전체 1163편의 논문을 대상으로 할 때는, 79%에서 99%사이의 논문이 자기 연구의 방법에 대한 분석이나 평가없이, 기존연구와 비교하여 일치하다는 것을 기술하는데 치중하고 있다고 추정할 수 있다.

자신의 연구결과와 기존연구결과와의 일치성(consistency)을 밝히는 것은, 변수들의 원인-결과 관계를 입증하는데 주로 쓰이는 기준중의 하나이며, 데이터의 과학적 신뢰성을 입증해주는 기준이 아니라는 것을 고려할 때, 총괄 및 고찰 부분에서 기존연구결과와의 일치성을 서술하는 것에 치중하는 것은 적절하지 않다. 같은 이유에서, 기초연구논문에서 연구결과를 설명하기 위하여 가능성이 있는 생물학적 기전(biological plausibility)을 언급하는데 치중하는 것도 원인-결과 관계 입증의 기준은 될 수 있어도⁽⁷⁾ 데이터의 신뢰성을 입증하는 기준은 되지 못하기 때문에 이 또한 적절하지 않다고 할 수 있다.

임상연구논문에서는 총괄 및 고찰부분에 교과서에 나오는 치료방법을 기술하는 경우가 많이 관찰되었다. 연구논문은 즉 원저(original article)는 어떤 새로운 사실을 발견하여 보고하는 것이 주목적임을 고려할때⁽⁸⁾, 새롭게 개발한 치료방법이 아니라 이미 널리 알려진 치료방법을 서술하는 것은 적절치 않으며, 굳이 서술하려면 다른 제목 밑에서 (임상가를 위한 특집이나 종설에서) 다루어져야 하는 것이 더욱 적절하다고 할 수 있다.

표7에서는 연구논문을 왜 독자들이 읽기를 기피하는가, 연구결론을 왜 임상에 적용하기가 어렵다고 하는가에 대한 답을 찾을 수 있다. 연구목적은 논문의 서론부분에 제시되므로 연구의 임상적 중요성이 분명하고 이해하기 쉽게 명시되어 있어야만 독자들로

표7. 임상적 가치성의 평가

연구목적의 임상적 중요성 및 연구결론의 임상 적용성이 제시되지 않은 논문 : 연구목적의 임상적 중요성 및 연구결론의 임상 적용성이 제시된 논문	비율 26 : 10 (72% : 28%)	95% 신뢰구간 [0.57 - 0.87]
---	------------------------------	---------------------------

하여금 읽고자 하는 동기를 부여할 수 있으며, 연구결론은 논문의 끝 부분에 서술되어서 연구결과를 임상에 어떻게 적용할 것인가 하는 방향을 제시해주어야 하는데, 대부분의 논문은 이 두가지 기능을 적절하게 하지 않고 있다는 것을 알 수 있다. 전체 1163편의 논문을 상대로 추정하면, 최소 57%에서 최대 87%까지의 논문이 독자들로 하여금 읽고 싶은 동기를 부여하지 못하고 있고, 설명 읽고나서도 연구결과를 임상적으로 적용하는데 도움을 주지 못하고 있다는 것을 나타내고 있다.

총괄 및 고찰

1. 표본의 크기, 선정방법 및 통계분석

이 연구에서는 사용된 표본은 2년동안 대한치과의 사회지에 게재된 연구논문 36편이었는데, 정식으로 난수표(table of random numbers)나 컴퓨터프로그램(computer generated random numbers)에 의한 확률표본(random sampling)에 의해서 형성되지는 않았다. 하지만, 전체 1163편의 연구논문중에서 확률표본을 이용하여 표본을 추출하는 방법을 사용하게되면 확률표본에 의해 선정된 논문들이 1950년이나 1960년에 발행된 경우에는 인터넷상에서 실제로 그 논문을 구하는 데 어려움을 예상할 수 있었다. 이런 연구방법의 어려움보다 더 중요한 사항은, 최근의 연구방법 및 주제분야를 분석 위해서는 최근에 쓰여진 논문을 대상으로 하는 것이 적절하다고 판단하였고 따라서 최근 2년동안의(본 연구는 2001년 4월부터 시작하였음) 연구논문을 연구자료로 선택하였다.

아울러, 표본을 통해서 전체를 추정하려는 연구의 목적을 달성하기 위해서는 모집단의 비율을 추정하는

데 필요한 전제조건 즉 표본수가 30편 이상되어야 하기 때문에 지난 2년동안의 연구논문중에서 데이터가 없거나 본문을 구할 수 없는 논문 4편을 제외한 36편 전체를 표본으로 한 것이다.

또한 표본선정과정에서 확률표본을 사용하려는 과학적인 목적은 선정된 표본이 모집단을 대표할 수 있도록 하기 위한 것임을 고려하면, 지난 2년동안의 논문을 표본으로 선정하는 연속표본(Consecutive Sample)을 사용한 것은, 현실적인 논문 확보의 여건과 최근의 논문의 경향을 대표한다는 점을 고려할 때 최선의 선택이었다⁽⁹⁾.

2. 분석기준 및 분석과정의 객관성

36편의 연구논문을 분석하는 기준은 Clinical Epidemiology에서 사용하는 과학적 연구방법의 원칙과 방법^(10,11,12) 및 통계학적 분석의 원칙과 방법⁽¹³⁾을 사용함으로써 분석기준의 객관성(objectivity)을 확립하였다. 분석자간의 차이(variation)를 줄이기 위해 1명의 분석자(본 연구의 연구자)가 3번에 걸쳐 각 분석기준에 의해서 분석하였으며 3번 모두 분석결과와의 차이는 없었다. 종합적으로 보면, 객관적기준과 동일측정자를 사용하여 반복분석하였기때문에 계통오차(bias)나 무작위오차(random error)는 연구결과에 심각한 영향을 미칠 정도로 크지는 않다고 볼 수 있다.

3. 분석과정에서 발견된 연구논문들의 문제점들

본 연구의 연구결과부분에서 보고 되지는 않았지만 논문들을 분석하는 과정에서 많이 발견되었던 문제점들은, 연구목적과 연구방법이 일치하지 않는 경우가 가장 많아서 연구목적은 “영향”을 알고자 하면서 연구방법은 상관관계의 규명으로 그치고 있었다⁽¹⁴⁾. 또한, 위약효과(placebo)를 배제시키기 위해서 반드시 있어야 하는 대조군을 사용하지 않은 상태에서 관찰한 어떤 치료방법이 효과가 있다고 보고하는 경우도 있었고⁽¹⁵⁾, 연구목적이 서론과 결론부분에서 서로 다르게 서술되고 있어서 혼동을 주는 논문도 있었으며, 데이터의 계측과정에서 발생하는 무작위 오차

(random measurement error)를 측정하기 위하여 표본의 추출과정에서 발생하는 무작위 오차(random sampling error)를 측정하는 방법을 사용하는 실수가 있는 논문도 있었으며⁽¹⁶⁾, 데이터에 근거하지 않고 개연성(plausibility)에 근거한 추측(guessing)만으로 서술을 하는 논문들도 있었다.

기초연구논문에서는 공통적으로 표본수의 크기가 불분명하여서 데이터의 안정성(stability)을 확신할 수가 없었고⁽¹⁷⁾, 동물실험에서 나온 연구결과를 어느 정도까지 인체에 일반화(extrapolation)시킬 수 있는가에 대한 서술이 없었다. 임상연구논문에서는 인간을 연구대상으로 하면서 윤리적 측면 즉 환자나 연구에 피연구자로 참가하는 사람들의 동의(informed consent)가 있었다는 것을 분명하게 밝힌 논문은 단 한편도 없었다. 또한, 혼란변수(confounding factor)의 개입가능성을 연구재료의 선정과정이나 데이터의 분석을 통하여 적절히 차단하거나 조정(adjustment)하지 않았고 아니면 적어도 개입가능성에 대해 인식하고 있는 논문도 적었다.

또한, 연구대상의 선정과정에서 처음에 연구대상으로 선정된 사람들중에서 연구에 참가한 사람들과 참가하지 않은 사람들의 차이를 조사하지 않은 경우도 많이 있었다. 기술적인(technical) 측면에서 보면, 데이터에 대한 설명과 표에 나와있는 데이터의 산술적 크기가 다른 논문, 영문으로 작성된 abstract와 한글로 작성된 본문내용이 서로 일치하지 않는 논문들은 독자들이 이해하는데 어려움을 느끼게 만들수 있는 가능성이 있었다.

결론 및 제언

연구결과와 총괄 및 고찰 부분을 종합하면, 다음과 같은 결론을 내릴 수 있다.

첫째, 연구재료, 연구방법, 통계분석방법, 총괄 및 고찰부분들이 과학적 연구원칙에 어긋나고 데이터에 오차가 많기 때문에, 연구결과를 신뢰하기 어려운 논문이 많다.

둘째, 연구주제의 임상적 중요성이 분명히 제시하지 않아서 독자들에게 논문들을 읽고자 하는 동기를 부여하지 못하고 있고, 연구결과를 임상에 어떻게 적용할 것인가에 대해 명확히 제시하지 않은 논문이 많아 독자들이 논문의 결론을 실제 임상에 활용하는데 어려움을 느낄 수 있다.

따라서, 이러한 문제점을 개선하기 위한 방안으로써 다음과 같은 두가지 제언을 한다.

첫째, 대한치과의사 협회지에 게재되는 연구논문들은 과학적 연구방법 및 통계분석방법에 더욱 충실하여야 하고 연구과제의 임상적 중요성 및 적용성을 더욱 명확하게 제시하여서 정확하고 유용한 지식을 제공하여야 한다.

둘째, 독자들에게 연구논문을 읽는 방법 및 통계학의 지식을 보수교육과정을 통해 제공하여, 게재된 연구논문을 과학적 원칙에 입각하여 비판적으로 예리하게 분석하면서 읽을 수 있고 정확하고 유용한 지식을 구별해서 습득할 수 있는 능력을 갖추도록 해주어야 한다.

참고 문헌

1. www.kda.or.kr 대한치과의사협회지 1999;10-12, 2000;3,8,9,10, 2001;1,4.
2. www.kda.or.kr 대한치과의사협회지 논문 목차 1999;12, 2000;1
3. Mendenhall, Beaver. Introduction to Probability and Statistics. 1994, p.285, p.514-515.
4. Wilson EB. An introduction to scientific research. New York, McGraw-Hill. 1952, p.169
5. Jamison JR. The use of inferential statistics in health and disease : A warning. S.Afr. Med. J. 1980;57:783-785.
6. Hulley SB, Cumming SR. Designing Clinical Research Williams & Wilkins, 1988. p.98-109.
7. Gordis L. Epidemiology. W.B.Saunders Company. 1996.
8. Hills AB. The reasons for writing. British Medical Journal. 1965:870.
9. Hulley SB, Cumming SR. Designing Clinical Research Williams & Wilkins. 1988, p.24-26
10. Feinstein AR. Clinical epidemiology : The Architecture of Clinical Research. W.B.Saunders Company. 1985.
11. Hennekens CH, Buring JE. Epidemiology in Medicine. Little, Brown and Company. 1987.
12. Wilson EB. An introduction to scientific research. New York. McGraw-Hill. 1952
13. Mendenhall, Beaver. Introduction to Probability and Statistics. 9th edition. 1994.
14. Phillips JL. How to think about statistics. 6th edition. 2000, p.150
15. Ederer F. Why do we need controls? Why do we need to randomize? American Journal of Ophthalmology. 1975:758-762.
16. Hulley SB, Cumming SR. Designing Clinical Research Williams & Wilkins, 1988. p.169.
17. Wilson EB. An introduction to scientific research. New York, McGraw-Hill. 1952, p.34.