

대한흉부외과학회지에 게재된 통계적 분석에 관한 고찰

송 현* · 박계현* · 김웅한** · 전태국**

=Abstract=

Present Statistical Status in Papers in the Korean Journal of Thoracic and Cardiovascular Surgery

Hyun Song, M.D.*, Kyeh Hyeon Park, M.D.*, Woong Han Kim, M.D.**, Tae Gook Jun, M.D.**

From January 1983 to December 1992, There were 1441 papers in the Korean Journal of Thoracic and Cardiovascular Surgery. Among these papers, 783 (54.3%) were original article or clinical analysis and 652 (45.2%) were case reports. A total of 319 papers contained some statistical analysis.

In 150 cases (47.0%) of these 319 papers, the statistical description was insufficient. Of the correctly described papers, 115 (68%) had more than one statistical error. Of course, in many cases the errors were not considered to be severe, but they were often sufficient to raise doubts about some inferences.

We suggest that authors should be more careful when they describe and apply statistical methods. If possible, authors should interpret results with statistical specialists. And we also suggest that our society have more extensive statistical refereeing system. This would at least prevent the worst errors from appearing in print. The last suggestion is elementary instruction in statistical methods during preclinical training.

(Korean J Thoracic Cardiovas Surg 1994;27:732-7)

Key words : 1. Statistics

2. Korean journal of thoracic and cardiovascular surgery

서 론

의학 분야의 연구가 신뢰할 만한 결론에 도달하기 위해서는 적절한 실험 계획, 자료 수집 및 자료 분석이 필요하다. 적절한 통계적 방법의 적용은 올바른 자료 분석에 있어서 필수적 요건이나, 저자나 감수자, 편집자가 모두 통계에 있어서 비전문인인 경우, 잘못된 통계 방법의 적용이나 분석으로 여러 가지 문제점을 야기시킬 수가 있다. 그

런 통계 분석의 오류가 환자의 치료와 치료 방법의 개발에 악영향을 미칠 수 있음은 자명한 사실이다. 흉부외과 분야가 눈부신 발전을 하고, 서로 많은 정보를 교환하게 된 이 시점에도 이와 같은 위험은 상존하고 있는 것이다.

이에 본 연구는 1983년 1월부터 1992년 12월까지 대한흉부외과 학회지에 발표된 논문 1441편을 대상으로, 각 논문이 통계학적 기법은 무엇을 사용하였고, 통계 방법의 기술은 어떤 식으로 하였으며, 그 오류는 무엇인가를 살펴보

* 국군 서울 지구병원 흉부외과

* Department of Thoracic and Cardiovascular Surgery, Seoul District Armed Forces General Hospital

** 서울대학병원 흉부외과

** Department of Thoracic and Cardiovascular Surgery, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

통신저자: 송 현, (110-200) 서울시 종로구 소격동 165, Tel. (02) 397-3838, Fax. (02) 397-3807

표 1. 논문 분포(1983년 1월부터 1992년 12월까지)

논문분류	수(N)	%
원 저	172	11.9
임상연구	611	42.4
증례보고	652	45.2
기 타	6	0.4
합 계	1441	100

아, 문제점을 파악하고, 이에 대한 향후 개선책과 지침을 제시하고자 하였다.

연구대상 및 방법

1. 연구 대상

연구 대상은 1983년 1월부터 1992년 12월 까지 대한흉부외과 학회지에 발표된 논문 1441 편을 대상으로 하였다. 이 중 증례보고 652 편(45.2%) 및 권두언, 종설, 기타 6편을 제외한 원저 172편(11.9%), 임상 연구 611 편(42.4%)의 총 783 편을 대상으로 하였다(표 1).

2. 연구방법

크게 두 가지의 관점에서 논문들을 살펴 보았는데, 그 하나는 통계 방법의 기술에 있어서 문제점은 없는가 하는 것이고, 또 하나는 통계 분석의 시행 시에 각 통계 기법의 오류에 관한 것으로 그 판단 기준은 다음과 같았다.

1) 통계 방법의 기술

① 연구 방법란에 저자들이 적용한 통계 분석 방법이 기술되어 있는가?

② 저자들이 기술한 통계 방법들이 실제로 그 논문에서 적용되었는가?

2) 통계 분석의 문제점에 관한 기준

이에 대하여는 국내 의학 통계서적¹⁻³⁾을 참고로 하여 기준으로 정하였는데, 각각의 통계 방법들에 있어서 잘못 적용되었다고 평가를 내린 기준은 다음과 같다.

① 평균치의 분석에 있어서의 문제점에 관한 기준

T-검정법의 경우에는 표본의 크기가 모수적 추론을 하기에는 적절하지 못한 경우, 분산이 지나치게 편이 되어 있어서 정규 분포를 가정할 수 없는데도 모수적 추론을 한 경우, 짝짓기(pairing)가 적절하지 못한 경우, 3군 이상에서의 비교를 2군씩 반복적으로 시행한 경우를 살펴 보았고, ANOVA 검정법의 경우는 사후 검정(중다 분석)을 적절히 사용하였나 하는 것을 살펴 보았다.

표 2. 통계 방법의 기술에 있어서의 문제점(n=319)

문제점의 분류	N	%
논문 내용에 통계 방법의 기술이 없이 통계적 의미만 부여한 논문	123	38.6
특정 통계 방법이 적용되었다고 기술되어 있으나 실제 적용되지 않은 논문	9	2.8
결과란 또는 고찰란에만 통계 방법이 기술되어 있는 논문	16	5.0
어느 항목이 어느 통계 분석을 하였는지 전혀 모르게 되어 있는 논문	2	0.6
연구 방법란에 통계기법을 정확히 기술한 논문(기술에 문제가 없는 논문)	169	53.0

② 비연속 변수의 분석에 있어서의 문제점에 관한 기준
비연속 변수를 연속 변수로 바꾸어 통계 처리한 논문과 Chi-square 검정법을 시행한 논문 중 어느 한 cell의 이론적으로 계산된 기대치가 5 미만에서 적용된 경우를 살펴 보았다.

③ 상관 및 회귀 분석에 있어서의 문제점에 관한 기준
상관 및 회귀 분석을 한 후에 이에 대한 통계적 검정이 이루어지지 않은 경우를 살펴 보았다.

④ 생존 자료 분석에 있어서의 문제점에 관한 기준
어떠한 생존 자료의 분석을 하였는지 언급이 없는 경우, 표본의 수가 적는데도 생명표 법을 이용하여 생존율을 구한 경우, 생존율을 구한 후 표준 편차 혹은 표준 오차의 언급이 없는 경우를 살펴 보았다.

연구결과

1983년 1월부터 1992년 12월 까지 10년간 대한흉부외과 학회지에 발표된 논문은 총 1441편 이었는데, 그 중 원저는 172편(11.9%), 임상 연구 611편(42.4%), 증례보고 652편(45.2%)이었고, 권두언 및 종설, 기타는 6편이었다(표 1). 원저 및 임상 연구 783편 중에서, 단순히 남자 몇 명, 여자 몇 명, 나이의 평균은 얼마 등 단순한 백분율 및 평균, 표준편차만을 기술한 논문은 464편으로 59.3%이었고, 통계적 접근은 319편(40.7%)에서 시행되었다.

1) 통계 방법의 기술에 있어서의 문제점(표 2)

통계적 접근을 시행한 논문 319편중, 통계 방법의 기술에 있어서, 논문 내용에 통계 방법의 기술이 전혀 없이 단지 p-value 및 통계적 의미를 부여한 논문은 123편(38.6%)이었으며, 연구 방법란에는 특정 통계 방법이 적용되었다

고 기술되어 있으나 실제 적용되지 않은 논문은 9편(2.8%), 연구 방법란에는 사용된 통계 방법이 기술되어 있지 않고, 결과란 또는 고찰란에서만 기술되어 있는 논문은 16편(5.0%)이었다. 또한 여러 종류의 통계 분석을 한 경우에 어느 항목이 어느 통계 분석을 하였는지 전혀 모르게 되어 있는 경우도 2편 있었다.

2) 통계 분석에 있어서의 문제점

통계적 접근을 시행한 논문 중, 기술상의 문제로 통계 방법의 오류를 검정할 수 없었던 125편을 제외하면, 통계 방법의 오류를 검정할 수 있었던 논문은 194편이었다.

1. T-test에서의 문제점

통계 방법의 오류를 검정할 수 있었던 논문 194편 중에서, T-test를 시행한 논문은 117편(60.3%) 이었는데, 분산의 차이가 있는데도 student T-test가 적용된 논문은 33편(28.2%) 이었으며, 짝짓기(pairing)가 적절하지 못한 논문은 21편(17.9%), 3군 이상에서 반복적으로 T-test를 시행한 경우는 45편(38.5%)이었다.

2. 분산 분석에서의 문제점

분산 분석을 시행한 논문은 총 24편(20.5%) 이었는데, 분산 분석을 시행한 후 중다비교(사후 검정)가 시행되지 않은 논문은 5편(20.8%), 분산 분석을 시행한 후 중다비교로써 T-검정법을 시행한 논문은 7편(29.2%), 분산 분석 없이 바로 중다분석이 시행된 경우는 3편(12.5%)이었다.

3. 비연속 변수에 있어서의 문제점

비연속 변수의 분석을 시행한 논문은 35편(18.0%) 있었는데, 이중 Chi-square 검정을 시행한 논문은 26편(74.3%), Fisher 직접 확률법을 시행한 논문은 11편(31.4%) 이었고, 2편에서는 Chi-square 검정과 Fisher 직접 확률법을 동시에 사용하였다. Chi-square 검정을 시행한 논문 26편 중, Cross-table에서 어느 한 cell의 이론적으로 계산된 기대치가 5미만에서 Chi-square 검정이 적용된 경우는 13편(50%), 대상군이 15이하에서 Chi-square 검정이 적용된 경우는 1편(3.8%) 이었다.

4. 상관 관계 및 회귀 분석에 있어서의 문제점

상관 관계를 구한 논문은 48편(24.7%) 있었고, 회귀식을 구한 논문은 31편(16.0%) 있었는데, 상관 및 회귀 분석을 시행한 후에 통계적 검정이 이루어지지 않은 경우가 11편(22.9%) 있었다.

5. 생존 자료 분석에 있어서의 문제점

생존 자료 분석을 시행한 논문은 72편 있었는데, 자료 분석의 방법에 대한 기술이 없는 경우가 28편(38.9%), 표본의 수가 적은데도 생평균 법을 이용하여 누적 생존율을 구한 경우는 11편(15.3%), 표준 오차(standard error)의 표현이 없는 경우가 26편(36.1%) 있었다.

고찰

의학의 연구 분야에서 연구 결과에 대한 해석과 그 지적 공유에 대한 의학 통계의 중요성은 두말할 필요가 없지만, 의학 분야에 종사하고 있는 임상 의사나 연구자 중 통계를 적절히 사용할 수 있는 숫자는 많지 않은 것이 현실이다. 결과적으로 의학 전문지를 읽는 독자들은 어떤 논문이 전문지에 발표되었을 때, 감수자나 편집자가 통계 분석을 포함하여 모든 측면을 세밀히 분석한 후에 게재한 것으로 생각하고 그 결론을 객관적인 타당성이 있는 것으로 간주하게 되지만, 통계 분석을 살펴 본 많은 연구자들은 사용된 통계 방법의 반수에서 그 분석이 적절하지 않음을 지속적으로 발표하였다⁶⁻⁸⁾. 국내에서도 통계의 관심도가 높아지면서 최근에는 국내 잡지에도 자체 내의 통계 분석에 관한 논문을 게재하게 되었는데, 역시 반수 이상에서 그 오류를 지적하고 있다^{9, 10)}.

본 연구에서 최근 10년간 대한흉부외과 학회지에 게재된 총 1441편의 논문 중에서 통계의 기법을 사용한 논문 319편을 살펴 본 결과, 연구 방법란에 통계 기법을 정확히 기술한 논문은 169편으로 53.0%였고, 이 169편의 논문 중에서 통계 기법상의 아무 문제도 없었던 논문은 54편으로 32.0%에 해당하였고, 통계 기법을 사용한 전체 논문으로 계산하여 보면, 16.9%밖에 안되는 수치였다. 전술한 다른 연구들에 비하면 좀 더 낮은 수치이지만, 이는 본 연구에서 정한 기준이 기존의 연구 기준에 비하여 관찰 항목이 더 많았기 때문에 발생한 연구 결과라 생각된다. 하지만 결코 16.9% 이외의 논문이 다 문제가 있는 논문이란 말은 아니다. 왜냐하면 다소간의 문제점이 있었다 하더라도 그러한 문제점 때문에 논문의 결과에 모두 심각한 영향을 미쳤다고는 할 수 없고, 또한 본 연구에서 정한 기준이 절대적일 수 없기 때문이다. 다만 적절한 통계 방법의 기술과 적용으로 보다 발전된 학회의 잡지를 가질 수 있으리라는 것이다.

통계 방법의 기술적인(descriptive) 측면이란, 어떤 면에

서는 논문 저자들의 성의하고 관계된 일이다. 하지만 표 2에서 보다시피 전체 논문 319편 중 169편(53.0%)만이 통계 방법의 기술에 있어서 문제가 없었다. 이는 아직도 흉부외과 학회 차원에서 통계 방법의 기술적 체계를 갖추지 못하고 있다는 것을 반영해 주는 것이다. 저자의 의견으로는 통계 방법을 기술할 때, 논문의 연구 방법란에서 연구자들이 적용한 통계 방법이 명시되어야 하며, 여러 종류의 통계 방법을 사용한 경우는 단순히 자신이 적용한 통계 방법을 나열하지 말고, 독자가 혼동되지 않도록 각 통계 방법의 기법을 어떤 결과에 적용하였는지 명시하여야 한다. 흔히 사용되지 않는 통계 방법이 적용될 때에는 그 통계 방법에 대한 간단한 설명을 가하는 것이 독자의 이해를 돕는데 도움이 되리라 생각된다. 또한 논문에서 결과란에 표(Table)를 제시할 경우는, 표 그 자체로써 독립적으로 설명할 수 있는 체계를 갖추어야 하기 때문에, 표에도 대상군의 숫자, 적용된 통계 방법, 통계적 검정 결과가 명시되는 것이 합리적이라 생각된다.

평균치의 비교는 임상의학 분야에서 가장 많이 사용되는 통계 방법인데, 크게 세 가지를 고려해야 한다. 첫째는 실험 대상군의 크기이고, 둘째는 관찰 대상군의 수이고, 셋째는 관찰하고자 하는 대상군의 관측이 서로 종속적이냐 독립적이냐 하는 것이다. 첫째로 실험 대상군의 크기는, 모수적 방법을 사용하여야 하느냐 비모수적 방법을 사용하여야 하는가를 결정하는 요소인데, 비모수적 통계 분석은 다음과 같은 경우에 적용해야 한다. 1) 수집한 자료가 명백하게 정규 분포를 따르지 않는 경우, 2) 자료의 관측 척도가 명칭 척도나 순위 척도인 경우, 3) 표본 수가 너무 적어 분포를 알 수 없는 경우로써 통상 한 집단의 표본 수가 10전 후 또는 두 집단인 경우는 합하여 20전 후일 때 사용하는데, 최소한 표본의 크기가 6이하인 경우는 비모수적 방법을 적용할 수 밖에 없다²⁾. 둘째로 관찰 대상군의 수에 따라서 T-검정법을 사용하여야 하는가, ANOVA 검정법을 사용하여야 하는가를 결정하여야 하는데, 3군 이상의 상황에서 두 군간의 차이를 검정할 때 T-검정법을 사용하면, 보정 없이 T-검정법이 반복적으로 사용되므로 α -error가 증가하여, 의미 없는 것을 의미 있다고 하는 결과를 초래하게 되므로 조심해야 한다. 따라서 대상군이 3군 이상인 경우에는 ANOVA test후에 사후 검정(다중 비교)로 Bonferroni correction, Turkey, Duncan, Neuman-Keuls, Scheffe, Dunnett 등을 적용해야 한다³⁾. 셋째로 평균치의 비교에서 흔히 발생하는 오류는, 관찰하고자 하는 대상군의 관측이 서로 종속적이냐 독립적이냐 하는 것에 대한 오류이다. 예를 들어 수술 전과 수술 후의 cardiac index의 측

정을 하였다고 가정하면, 수술 후에 cardiac index가 정말 증가되었는가를 알기 위해서는 짝지은 통계, 즉 paired T-test를 하여야 하는 것인데, 그냥 T-test(unpaired T-test)를 시행하는 경우이다. 이 경우 수술 전의 cardiac index는 수술 후 수치에 영향을 주기 때문에(종속적인 관계) paired T-test를 시행하여야 한다. 만약 종속적인 관측 치에 대하여 unpaired T-test를 시행하면 통계적으로 의미 있는 결과를 의미 없다고 하게 되는 β -error가 증가되므로 주의해야 한다^{1,2,5)}.

평균치의 비교 중 최근에 점점 그 이용이 증가되고 있는 분산 분석(Analysis of variance, ANOVA)은 3군 이상의 평균치를 비교하는 통계 기법인데, 본 연구에 의하면 주로 사후 검정과 관련되어 그 오류가 관찰되었다. 분산 분석을 시행 후 의미 없는 결과가 나왔을 때는 일반적으로 더 이상의 통계 분석은 필요치 않으나, 통계적으로 의미 있는 결과가 나왔을 때는 여러 군중에서 어느 두 군간의 차이에 의해서 전체적으로 의미가 있게 나왔는지 검정하게 되는데, 이 때에 적용하게 되는 방법이 사후 검정 혹은 중다 비교이다. 종류로는 전술한 바와 같이 Bonferroni correction, Turkey, Duncan, Neuman-Keuls, Scheffe, Dunnett 등이 있다. 대한흉부외과 학회지에는 10년간 총 24편의 분산 분석을 이용한 논문이 있었는데, 분산 분석을 시행한 후 중다비교(사후 검정)가 시행되지 않은 논문은 5편(20.8%), 분산 분석을 시행한 후 중다비교로써 T-검정법을 시행한 논문은 7편(29.2%), 분산 분석 없이 바로 중다분석이 시행된 경우는 3편(12.5%)으로 문제가 없는 논문은 10편으로 41.6%였다.

비연속 변수의 통계적 분석은 연속적이지 못한 변수, 즉 빈도(frequency) 또는 비율(ratio)을 비교하는 방법인데, 성별, 환자군/대조군, 수술 후 생존/사망 등 둘로 갈라지는 변수가 좋은 예이며, ABO혈액형, 농도를 달리한 약물 투여군, 등급으로 나눈 병리학적 분류, 항체 titer 등 세 가지 이상으로 갈라지는 양적인 다분성 변수에 대한 통계학적 분석이다. 물론 원칙적으로는 연속적 변수라 하더라도 분석을 위해 분류하는 과정에서 얼마든지 비연속 변수로 변환될 수 있다. 대표적인 분석 종류로는 Chi-square test, Fisher's test, Score test for trend, Ridit test 등이 있다.

Chi-square 검정법이 적절히 적용하려면 몇 가지의 기본 조건이 만족되어야 한다. 첫째는 대상군의 숫자가 커야 하는데 대개 50이상이 요구된다⁴⁾. 자체 대상군의 숫자가 15이하인 경우에는 Chi-square 분포에 의한 오차가 커지기 때문에 Fisher's test에 의하여 통계 검정을 하여야 한다. 둘째는 Cross tabulation에서 한 cell의 이론적으로 계산된 기

대치가 최소한 5 이상은 되어야 하는데, 5 미만인 경우는 Yates의 수정법에 의한 Chi-square 검정법을 적용하던가, Fisher's test를 사용하여야 한다. 셋째는 관측치가 서로 독립적인가 종속적인가에 대한 고려가 있어야 하는데, 독립적인 경우에는 Chi-square 검정법을 적용하나, 동일한 대상에 대하여 특정 처치 전의 검사 결과와 처치 후의 검사 결과를 비교하는 것과 같은 종속적인 경우에는 McNemar Chi-square test를 적용해야 한다²⁾. 본 연구에서는 Chi-square 검정법을 적용한 논문 중에서 기대치가 5미만인 경우만 살펴 보았는데, Chi-square 검정법이 적용된 26편 중에서 13편에서 그 기대치가 5미만인 경우인데 적용하였다. 7편에서는 비연속 변수를 연속 변수로 바꾸어 통계 분석을 시행하였는데, 예를 들면 functional class 같은 비연속 변수를 연속 변수로 바꾸어(수치화 하여) 비교한 것으로, 이것도 문제가 있는 통계 처리라 하겠다. 이런 functional class같은 정도(grade)가 매겨지는 경우는 다분히 주관적 판단에 근거하는 자료이며, 만약 1, 2, 3, 4로 계량화 한다고 하여도 1과 2 혹은 2와 3간의 등 간격이라고 단언할 수 없는 처지에 놓이게 된다. 따라서 이때는 각 순위를 계량화 한다는 생각은 버리고, 대신 관찰된 자료만을 가지고 분석하는 Ridit(Relative to an identified distribution) 분석법을 사용하여야 한다³⁾.

자료 분석에 있어서 흔히 두 변수간의 관계에 대하여 알고 싶을 때가 있다. 예를 들어 혈압과 나이의 관계, 몸무게와 키, 주사약품 농도와 심박동수 같은 변수들간의 관계 정도를 분석하는 것이 상관 분석과 회귀 분석인데, 두 통계 기법이 어느 정도 관련은 있지만, 실제로 사용되는 목적은 서로 다르다. 회귀 분석은 변수들간의 관계를 파악하는데 유용하며, 이 기법을 사용하면 어떤 변수를 통하여 다른 변수의 값을 예측(predict) 또는 추정(estimate)할 수 있다. 한편 상관 분석은 변수들간 관계의 강도를 측정하는 것이다³⁾. 즉 상관 분석은 두 변수간의 선후 관계가 분명하지 않을 때 사용하는데 비하여 회귀 분석은 두 변수간의 선후 관계가 분명한 경우에 사용하게 된다³⁾. 상관 분석을 하는데 있어서는 모수적 방법과 비모수적 방법이 있는데, 모수적 방법으로는 대표적으로 Pearson 상관계수가 있고, 비모수적 상관계수로는 Spearman 상관계수, Kendall의 Tau-b 상관계수, Hoffding Dependence 상관계수가 있다. 어떠한 종류의 상관계수를 구하던지 상관계수 자체가 통계적 검정력을 가지고 있지 않기 때문에 상관계수에 대한 유의성 검정을 하여야 한다. 10년 동안 대한흉부외과 학회지에 발표된 논문들 중에서 상관관계나 회귀분석을 시행한 논문은 48편이 있었는데, 힘들게 통계 처리를 한 후에

P-value에 대한 언급이 없는 논문이 11편으로 23%에서 발견되었다.

생존 자료의 분석은 valve failure 등의 complication 및 cancer patient의 survival 때문에 흉부외과 영역에서 많이 사용되는 통계 영역이라 할 수 있다. 생존율의 산출 방법은 크게 실측 생존율(actual survival rate)과 상대 또는 보정 생존율(adjusted survival rate)로 나눌 수 있는데, 실측 생존율은 다시 직접법에 의한 생존율과 누적법에 의한 생존율로 나눌 수 있다. 대부분의 경우에 있어서 생존율은 누적 생존율로 산출되는데, 누적 생존율이란 구간 또는 순간 생존율을 구하고 이들을 누적하여 산출하는 방법을 의미한다. 일정 함수(예: 지수 함수)를 이용하는 모수 방법(parametric method)과, 생명표 법(life-table method), product-limit 법, Chiang's 법 등의 비모수 방법(nonparametric method)가 있다³⁾. 이중 생명표 법(life-table method, Cutler-Ederer method)은 그 중에서도 가장 많이 사용되는 방법인데, 일반적으로 한 군의 표본 수가 50은 넘어야 한다. 따라서 표본의 크기가 작을 때는 Kaplan-Meier(Product-limit)법을 사용해야 한다^{2, 4)}. 대한흉부외과학회지에는 통계 기법을 사용한 총 319편의 논문 중 생존 자료의 분석은 72편(22.6%)에서 시행되었는데, 분석에 있어서 문제가 없었던 논문은 24편으로 1/3 이었고, 나머지는 생존율을 구한 방법의 기술이 없거나, standard error의 언급이 없거나, 관찰 수가 적어서 Kaplan-Meier의 누적 생존율을 구하여야 하는데도 life table method를 사용하여 문제가 있는 논문이었다.

누적 생존율이나 누적 생존 곡선이 치료군 또는 연구하고자 하는 예후 인자에 따라 차이가 있는지를 검정하기 위해서는 두 생존 곡선의 비교를 시행하여야 하는데, 그 종류로는 Greenwood 근사식(Peto의 간편식)법, Mantel-Haenszel법, Log rank 법, Wilcoxon rank 법, Likelihood ratio 법 등이 있다. Greenwood 근사식(Peto의 간편식)법은 일정 시점까지의 혹은 특정 관찰구간에서의 두 생존율 차이에 대한 유의성 검정은 가능한데, 전체 관찰기간을 통한 생존 양상에 차이에 대한 유의성 검정은 불가능하다. Mantel-Haenszel법은 censored cases를 검정 과정에서 모두 제외시키기 때문에 증례수가 적어져 소위 통계적 검정력이 다른 방법에 비해 낮은 단점이 있다. 그 외 Log rank 법은 생존 기간이 긴 자료, Wilcoxon rank 법은 생존 기간이 짧은 자료에 적절하며, Likelihood ratio 법은 생존 기간의 분포가 exponential 분포를 따를 때 사용이 가능하다^{2, 5)}.

결론적으로 본 연구에서 관찰된 10년간 대한흉부외과학회지에 게재된 총 1441편의 논문 중에서 통계의 기법을 사

용한 논문은 319편이었는데, 그 중에서 연구 방법란에 통계 기법을 정확히 기술한 논문은 169편으로 53.0%였고, 이 169편의 논문 중에서 통계 기법상의 아무 문제도 없었던 논문은 54편으로 32.0%에 해당하였고, 통계 기법을 사용한 전체 논문으로 계산하여 보면, 16.9%밖에 안되는 수치로, 이것은 흉부외과 영역에서의 정확한 통계 방법의 적용에 얼마나 많은 사람들이 무관심 했나 하는 것을 보여주는 소견이라 할 수 있겠다.

이러한 문제점을 해결하는 방법으로써, 본 저자는 다음의 몇 가지 개선책을 제시하여 보았다.

첫째는 본인의 노력이다. 연구 방법란에 통계 방법을 정확히 기술하고, 잘 모르는 것은 통계학자와 상의를 하고, 결과의 해석과 기술하는 방법까지도 조언을 받고, 통계학 책을 참고하는 등의 본인의 충분한 검토와 노력이다. 그럼으로써 연구 결과란에만 의미를 부여하는 등의 잘못은 손쉽게 없어지리라 본다.

둘째는 학회 차원에서의 논문 심사 강화이다. 우선 논문의 투고 요령에 통계 방법의 기술 및 적용에 대한 지침을 마련하고, 가능하다면 편집위원 중에 통계 전문가를 두어, 제출된 논문에 대하여 통계 방법에 대한 검토를 시행하는 것이다.

셋째로는 통계학에 대한 교육의 강화로, 현실적으로는 어려움이 많으리라 생각되지만, 전공의 교육과정, 또는 대

학원 과정에서의 교육 강화를 생각해 볼 수 있다.

이상으로 1983년부터 1992년까지 10년간 대한흉부외과학회지에 발표된 논문들을 대상으로 통계 방법의 기술과 적용에 있어서의 오류를 살펴 보았고, 그 개선책을 제안하여 보았다.

References

1. 고응린. 계량의학통론. 서울:신광출판사. 1982
2. 안운옥. 실용의학 통계론. 1판. 서울:서울대학교 출판부. 1990
3. 이승우. 통계학의 이해. 개정판. 서울:자유아카데미. 1993
4. 박순영. 기초의학 통계학. 초판. 서울:이공도서 출판사. 1976
5. 서울대학교 의과대학 예방의학교실. PC-SAS 공개강좌. 개정판. 서울. 1992
6. 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
7. Schor S, Karten I. *Statistical evaluation of medical journal manuscripts.* JAMA 1966;195:1123-8
8. Glantz SA. *Biostatistics: How to detect, correct and prevent errors in the medical literature.* Circulation 1980;61:1-7
9. 조수철, 안운옥. 신경정신의학에 발표된 논문들의 통계방법의 기술 및 적용 현황:문제점 및 개선책. 신경정신의학 1991;30:462-73
10. 고흥, 박일용, 김광우, 함병문, 최익현. 대한마취과학회지에 게재된 논문의 통계적 분석에 관한 고찰(1981년-1990년). 대한마취과학회지 1993;26:22-7