

의학 연구자료 분석과 통계적 기법

안 윤 옥¹⁾

1. 서론

醫學(medicine)은 '건강의 유지와 질병의 예방, 그리고 완화 또는 치료에 관한 과학이며 예술이다' (the science and art dealing with the maintenance of health, and prevention, alleviation or cure of disease) 라고 Webster 사전에 정의되어 있다. 즉, 學問研究와 醫術의 시행이 그 기본 구성요소인 것이다. 학문연구를 통하여 얻어진 지식을 바탕으로 의술이 행하여 지는데 그 시술의 대상은 바로 사람이기 때문에 의학에서의 학문연구는 科學性과 함께 倫理性이 매우 중요한 고려사항이 된다. 일반적으로 과학적이라 함은 연구의 계획에서부터 수행, 결과분석 및 결론도출에 이르는 모든 연구과정에 객관적이며 타당한 방법이 적용되었음을 뜻한다. 한편 윤리적이란 함은 의학적 지식이나 시술이 인간에게 이득(benefit)이 되어야 하며 사회정의(justice)에 합당하고, 자유의지(autonomy)가 보장되는 것으로 이해되고 있는 바, 연구방법 및 과정에서는 물론이고 연구결과와 해석이나 판단에 있어서도 윤리적 관점은 비중 높게 반영된다. 예를 들어 어떤 치료방법이 효과가 있는데 있다고 하는 것은 분명히 비윤리적이지만 반대로 효과가 있는데 없다는 결론에 이르는 것도 또한 비윤리적인 것이다.¹⁻²⁾

연구자료를 분석하는 가장 주된 목적은 연구자가 주장하고자 하는 연구결론에 대한 각종 증거들을 요약하는데 있다³⁾. 따라서 어떻게 분석할 것인가 하는 질문은 바로 주장하고자 하는 내용이 무엇인가 하는 질문이 된다. 바꾸어 말하면 연구목적이 무엇인가에 따라 자료분석 내용과 방법이 달라진다는 것이다. Yale대학의 A. Feinstein⁴⁾은 의학연구를 연구의 일반목적에 따라 記述的(descriptive)연구, 因果性(cause-effect)연구, 그리고 過程(process)연구로 구분하고 있고, 보다 구체적 연구목표에 따라서는 9가지로 분류하고 있다. 1) 생리적이전 연구, 2) 병태생리연구, 3) 발병요인 연구, 4) 정상치역 연구, 5) 검사측정방법 연구, 6) 진단능력평가 연구, 7) 질병에후인자 연구, 8) 약물대사 및 약동학적 연구, 9) 약물안정성 및 효능 연구 등이다. 요컨대, 의학연구자료의 분석에서는 이러한 연구목적에 따라서 거기에 합당한 기법이 다양하게 적용되고 있다. 그러나 최근 필자등의 조사보고⁵⁻⁶⁾에서 보면 1980년대에 우리 나라에서 발표된 의학연구논문의 대부분은 아직도 적지 않은 방법론상의 오류가 지적되고 있는 실정이다.

본 小考에서는 의학연구자료가 연구목적에 따라 어떠한 분석기법, 특히 통계적 기법이 적용되고 있는가를 나열, 소개하고자 한다. 아울러 국내 의학논문에서 흔히 나타나는 통계처리의 잘못을 Gore⁷⁾의 분류를 근간으로 하여 그 유형과 빈도를 제시한다.

2. 記述的(descriptive) 연구에서의 통계기법

특정 건강-질병현상의 빈도 혹은 분포양상을 파악하는 연구는 가장 초보적인 記述的 연구이다. 예를 들어 어떤 집단의 신장, 체중과 같은 생리적 상태의 분포는 물론 고혈압과 같은 질병상태의 발생수준이나 유병자율, 또는 사망수준 등을 파악하는 자료가 그 대표적이다. 일반적으로 대표값과 산포도(또는 신뢰구간)로서 연구자료가 요약되지만 대부분에 있어서는 다른(집단, 지역 또는 시점) 자료와의 比較分析이 이루어진다.

1) (110-799) 서울시 종로구 연건동 28 서울대학교 의과대학 예방의학교실 교수

측정치의 척도(scale)나 통계처리 대상자료의 형태에 따라 적용하는 통계기법이 다르다. 특히 의학연구자료에서는 nominal과 ordinal scale이 자주 사용된다(interval 내지는 ratio scale도 물론 사용되지만). 또한 분석대상 자료의 형태로는 ratio보다 rate나 proportion이 더 많이 사용된다.

(예 1)

간암환자(N)에서 B형간염 바이러스 s항원 검사를 하여 양성자가 a 명이라는 자료를 얻는다.

===> measurement는 nominal scale 이며
분석대상은 a/N의 proportion이 된다.

===> Binomial distribution or asymptotic χ^2 function 적용*

*보다 나은 approximation을 위하여 a/N를 arcsin square root로 variance-stabilizing transformation한다.

(예 2)

N 명의 주민이 사는 지역에서 1년 동안에 a 명이 위암으로 사망하였다. 년초인구는 N_0 , 년말인구는 N_t 이다.

===> measurement는 nominal scale 이며
분석대상은 $a/((N_0 + N_t)/2)$ 의 rate가 된다.

===> Poisson distribution or asymptotic χ^2 function 적용*

全圖의 위암사망 수준(위암사망율 m)과 비교하고자 한다

===> Standardized Mortality Ratio(SMR)를 구하여

$$SMR = (\text{No. of Observed deaths}) / (\text{No. of Expected deaths})$$

$$= a / (m \cdot (N_0 + N_t) / 2)$$

Poisson distribution을 적용한다

*보다 나은 approximation을 위하여 rate를 square root로 variance-stabilizing transformation한다.

(예 3)

N 명의 위암환자에 대하여 처음 진단시의 병기(stage)별로 구분한 자료는 아래와 같다.

초기: a, 1기: b, 2기: c, 3기: d, (a+b+c+d= N)

===> measurement는 ordinal scale 이며 분석대상은 a/N, b/N, c/N, d/N 의 proportion이다.

10년전의 자료 (N_1, a_1, b_1, c_1, d_1)과 비교하고자 한다

===> χ^2 distribution을 적용하기도 하고 RIDIT(relative to an identified distribution) 분석을 적용하기도 한다.

(예 4)

특정 정신질환의 진단을 위하여 임상적인 증상과 소견을 종합하여 일정 기준을 만들고 이를 시험적으로 적용하였다. 같은 대상에게 정신과 의사 2인이 독립적으로 적용해 보도록 하였고 그 중 1명의 의사에게는 시간 간격을 두고 2번 반복(test and retest)토록 하였다.

의사 B의 결과	의사 A의 결과		의사 A의 두번째 결과	첫번째 결과	
	질병(유)	질병(무)		유	무
질병(유)	a	b	질병(유)	a'	b'
질병(무)	c	d	질병(무)	c'	d'

===> 임상기준에 의한 진단방법의 신뢰성(reliability)을 분석 하고자 한다.

===> 일치도(indices of agreement, a와 d, a'와 d')를 분석한다. kappa statistics를 많이 적용한다.

3. 原因性(Risk factors) 연구에서의 통계기법

질병발생 원인을 규명하고자 하는 연구는 질병 개개의 病因論에 따라 접근방법 내지는 연구방법 및 형태가 다르다. 單一原因(또는 필요원인)이 병인론으로 되어 있는 질병(예: 각종 감염성 질환)에서는 주로 실험동물이나 세포를 이용하는 실험실적 연구를 수행하는데 연구자료의 해석은 소위 決定論的(deterministic) 입장에서 이루어지기 때문에 통계적 기법의 적용은 드물다. 그러나 한편 多原因 병인론으로 믿어지는 질환(예: 통상 성인병이라고 불리우는 암, 고혈압, 심장병 등의 만성퇴행성 질환)에 대하여는 그 원인성을 確率論的(probabilistic)으로 추론할 수 밖에 없다. 인구집단을 직접 대상으로 하여 수행하는 疫學的 연구자료나 실험동물등을 이용하는 실험실적 연구자료도 이 경우에는 통계기법의 적용이 필수적이 된다.

가상원인(X)과 특정질병(Y)이 인과성이 있다는 연구결과에 도달하려면 현실적으로 먼저 X와 Y간에 통계적인 相關性(Statistical association, Covariation)이 제시되어야 한다. 따라서 연구계획 단계에서부터 Statistical validity(alpha or beta error)가 있도록 통계적인 고려를 하게 된다. 또한 인과성의 추론을 위하여 연구설계상 비교집단(comparison group)을 연구집단(study group)과 함께 설정하여 자료를 수집, 분석하게 된다. 예를 들어 X요인이 작용하였던 집단(exposed group)과 없었던 집단(unexposed group), 또는 Y 질병이 있는 집단(case group)과 아닌 집단(control group)으로 조사대상을 설정하게 된다. 요즈음에는 이러한 비교집단을 개별 짝짓기(matching)로 선정하는데 선정된 개개의 연구대상(study case)에 대하여 攪亂변수(confounder)가 동일한 대조자를 뽑아 matched pair를 구성하게 한다.

X와 Y와의 相關 정도를 나타내는 지표(measures of association)를 산출하여 이를 분석하게 된다. 相關도 지표로는 difference measure와 ratio measure가 있는데 ratio measure에는 rate ratio와 odds ratio가 있다. 이를 요약하면 아래에서와 같다.

가상원인 여부	No.	관찰총기간	질병발생수	발생율
가상원인(무)	N_0	T_0	a	$r_0 = a/T_0$
가상원인(유)	N_1	T_1	b	$r_1 = b/T_1$

====> rate difference (RD) : $r_1 - r_0$
 rate ratio (RR) : r_1 / r_0
 odds ratio (OR) : $(b / (T_1 - b)) / (a / (T_0 - a))$
 $= r_1(1 - r_0) / r_0(1 - r_1)$

연구집단 구분	No.	과거 가상적 요인에의 폭로자	폭로 odds
질병자 집단	N_1	b	$b / (N_1 - b)$
대조자 집단	N_0	a	$a / (N_0 - a)$

====> odds ratio (OR) : $(b / (N_1 - b)) / (a / (N_0 - a))$

관련도 지표에 대한 통계적 분석기법의 적용은 연구대상 선정에 관련한 sampling 방법과 그 규모에 따라 다르다. hypergeometric distribution을 이용하는 Fisher's exact method는 주로 Small sample size에서 적용된다.

연구대상 집단(보통 2이상)의 선정이 집단별로 이루어 졌는가(grouped data), 또는 짝짓기로 이루어 졌는가(matched data)에 따라 분석기법은 전혀 다르다. 또한 measures of association의 형태 (RD, RR, or OR)에 따라서도 분석기법이 다르다.

3.1 RD의 분석기법

Null χ^2 function을 적용하는 경우와 Nonnull(basic) χ^2 function을 적용하는 경우로 구분하여 생각할 수 있다. 즉 parameter RD=0을 Expectation으로 하는 경우와 parameter RD값을 특별히 지정하지 않는 경우이다. RD=0인 상황에서는 variance estimator가 쉽게 추정될 수 있으나 Nonnull에서는 매우 복잡하게 추정된다. RD에 대한 asymptotic confidence bounds를 산출할 때는 basic χ^2 function이 적용된다.

가상적 원인(X) 이외의 제3 요인(F,예: 연령)의 간섭을 배제하기 위하여는 F변수를 여러 층(Stratum)으로 구분한 층화분석을 하는데, 짝짓기(F변수가 같도록 한 것임)로 수집된 자료(matched data)는 개개의 matched pair가 개개의 stratum에 해당하기 때문에 pair의 수만큼의 strata가 있게 되고 개개 stratum의 weight는 모두 동일한 특수한 경우의 층화분석에 해당하는 것이다.

Regression analysis의 경우에서 RD가 분석 대상인 경우는 General linear model이 선호된다. 왜냐하면 GLM에서의 regression coefficient는 RD의 estimator가 된다.

3.2 RR의 분석기법

r_1 의 분포와 r_0 의 분포를 이용하는 경우와 $\log(RR)$ 의 분포를 이용하는 방법이 적용된다. 즉,

basic χ^2 function으로;

(1) $\chi^2 = (r_1 - (RR)r_0)^2 / (V_{r_1/RR} + (RR)^2 V_{r_0/RR})$ 와

(2) $\chi^2 = (\log(r_1) - \log(RR))^2 / (V_{\log(r_1)/RR})$ 을 적용한다.

; $V_{\cdot/\cdot} = \text{Conditional Variance}$

상기 두 방법의 variance estimator를 정하기 위하여 (1)에서는 Likelihood equation의 답을 구

하게 되고, (2)에서는 Taylor series approximation을 하게된다. 그러나 parameter RR 값에서 $\log(rr)$ 이 지니는 분산은 $V_{\log(rr)}$ 로 대응하게 되고 이는 간편법으로 proportion type일 때와 rate type인 경우 다음과 같이 추정 된다.

$$V_{\log(rr)} = \left\{ \frac{(1-r_1)}{C_1} \right\} + \left\{ \frac{(1-r_0)}{C_0} \right\} \Rightarrow \text{proportion type}$$

$$= \left(\frac{1}{T_1} \right) + \left(\frac{1}{T_0} \right) \Rightarrow \text{rate type}$$

; C = 각 집단에서의 발생자 수
T = 각 집단에서의 총 관찰기간 수

regression analysis에서는 logistic regression model이 선호된다. logistic regression coefficient는 OR의 estimator인데 질병발생율(rate)이 낮으면 RR과 OR은 매우 근접한다. Matched data에서는 필히 stratified(or matched) analysis를 한다.

3.3 OR의 분석기법

RR의 분석기법과 동일하다. 단지 $RR = R_1/R_0$ 인데 반하여 $OR = R_1(1-R_0) / R_0(1-R_1)$ 이므로 RR에서의 (1)식은 약간 변형된다. (2)식은 statistic $rr \rightarrow or$, parameter $RR \rightarrow OR$ 이 된다. 간편법에 의한 $V_{\log(or)}$ 은 다음과 같다.

$$V_{\log(or)} = \left(\frac{1}{C_1} \right) + \left(\frac{1}{(N_1 - C_1)} \right) + \left(\frac{1}{C_0} \right) + \left(\frac{1}{(N_0 - C_0)} \right)$$

regression analysis에서는 logistic regression model이 선호된다. logistic regression coefficient는 OR의 estimator이며 질병발생율(rate)이 낮으면 RR과 거의 같은 값을 보인다. Matched data에서는 필히 stratified(or matched) analysis를 한다.

4. 치료효과 및 질병예후인자 연구에서의 통계기법

치료효과 및 질병예후인자에 관한 연구는 의학연구의 독특한 영역이다. 앞서 언급한 바와 같이 과학성과 윤리성이 엄격하게 요구되는 분야이다. 흔히 임상시험(Clinical trial)이라고 불리는 이 연구는 연구계획 단계에서부터 통계적 기법의 적용이 절실하게 요구된다. 사람을 직접 대상으로 하여 일종의 실험을 수행하는 것이기 때문에 연구의 시행착오는 용납되지 않는다. Statistical validity를 최대화 하기 위한 연구설계가 필요하며 혹시 시험대상에게 해(harm)가 될지도 모르기 때문에 Minimum sample size로 타당한 conclusion을 얻어야 한다. 다른 연구 계획에서와 같이 study size가 多多益善은 아니다. 수종의 치료방법을 비교하는 연구에서는 시험대상자에게 어떤 치료법을 배정할 것인지 에서도 윤리성이 고려되어야 한다. 통상 random allocation의 방법을 적용한다. 또한 시험연구 도중이라도 치료의 효과가 통계적으로 명백히 기대치 이하라는 결정이 난다면 연구는 중도에 포기하여야 한다.

수집자료의 분석에서는 χ^2 function이 적용되기도 하나 거의 대부분에 있어서는 Survival function을 적용한다.

5. Nonparametric methods적용의 증가

의학연구의 대부분은 특정한 상태의 사람(특정 질병자 또는 특정원인에 폭로된자 등등)이나 실험동물을 대상으로 하여 연구자료가 수집된다. 따라서 이론적으로 합당한 규모의 연구대상을 확보하기가 현실적으로는 매우 힘들다. 또한 윤리적인 측면에서 연구결론을 미리 예측해 보는 것이 필요한 경우도 많다. 또 한편으로는 결과관찰의 측도가 nominal 또는 ordinal인 경우가 많고 이들 측정치의 분포가 定形的(parametric)이지 않은 경우도 허다하다. 이러한 연구자료의 특성은 nonparametric statistical methods의 적용에 아주 적합하다.

6. 국내 의학논문에서의 통계기법의 적용 현황

1980년 - 1989년 10년 동안에 '대한의학협회지'에 게재된 '원저' 논문 467편중 명백히 '사례조사'나 '종설'등에 해당하는 85편을 제외한 나머지 382편에 대하여 통계기법 적용등을 포함한 방법론적 타당성을 평가하는 조사연구를 시행한 바 있다(6). 통계처리기법의 타당성을 평가한 결과를 전제한다. 통계적인 오류를 점검하는 기준은 Gore(7)의 분류를 근간으로 하여 세부 사항을 수정, 보완하였다.

Table. Number of Articles with Statistical Error(s) according to the Error Category (Medical articles in JKMA, 1980-1989)

Error category	No. of articles applicable for category	with Error(s)	
		No.	%
1. Error of Omission			
1) incomplete description of basic data	249	48	19.2
2) statistical test performed not yet defined	221	145	58.2
3) incomplete description of power or confidence interval	221	203	91.9
2. Error of Commission			
1) inadequate description of measures of central tendency or dispersion	201	56	27.9
2) incorrect analysis	104	50	48.1
3) multiplicity on hypothesis testing	221	144	65.2
4) unwarranted conclusion	297	155	52.5

참고문헌

- [1] 안윤옥, 이형기 (1990), "의학에서의 연구방법론," 한국역학회지 12(2):107-14.
- [2] 안윤옥 (1989), "임상시험의 설계," 한국역학회지 11(1):1-17.
- [3] 안윤옥 (1990), 실용 의학통계론. 서울대학교출판부, 서울
- [4] Feinstein A. R. (1985) *Clinical Epidemiology. The architecture of clinical research*. W.B.Saunders Company, Philadelphia.
- [5] 안윤옥, 이형기 (1991), "의학연구논문의 방법론 및 통계처리기법의 타당성 평가를 위한 점검표 개발," 한국의학교육 3(1):19-35.
- [6] 이형기, 안윤옥 (1991), "1980년대에 발표된 국내 의학연구논문의 방법론 및 통계처리기법의 타당성에 관한 평가연구," 한국의학교육 3(1):52-69
- [7] Gore S. M., Jones I. G. and Rytter E. C., (1977), "Misuse of statistical methods: critical assessment of articles in BMJ from January to March 1976," *British Medical Journal* 8 p.p. 85-87, 1977