

Logistic Analysis를 이용하여 분석한 인공유산이 속발성불임에 미치는 영향

가톨릭 의과대학 예방의학교실
<지도교수: 조 규 상 교수>

이 원 철

= Abstract =

The Study of the Influence of Induced Abortion on Secondary Infertility analyzed by Logistic Regression

Won Chul Lee, M.D.

Department of Preventive Medicine & Biostatistics, Catholic Medical College
(Directed by Professor Kyu Sang Cho, M.D.)

The methods controlling the confounding factors were discussed using the data of secondary infertility with induced abortion. Mantel-Haenszel method and logistic model were applied in the analysis to find out which factors were confounding and/or effect modification variables.

In the logistic analysis, the main effect of induced abortion, spontaneous abortion, age and interaction effect between induced abortion and spontaneous abortion were chosen as independent variables being regressed into logistic functions. Spontaneous abortion was interpreted as a potential confounder and at the same time potential effect modifier and age was interpreted as potential confounder. Spontaneous abortion was shown to be more important influencing factor than age to the secondary infertility.

In the course of logistic analysis, the problem of parameter estimation and hypothesis testing, assessing the fitness of a model, and selection of the best model were briefly explained. For the program of logistic model, FUNCAT Procedure of SAS package was chosen.

I. 서 론

종속변수(dependent variable)가 범주적(categorical)인 성격을 띄우는 자료를 분석하는 경우 교란인자(confounding factor)의 영향을 제거하기 위하여 사용되는 통계적기법으로서 표준화(standardization)와 층별(stratification), log-linear analysis, logistic analysis 등을 들 수 있다(Anderson et al, 1980)¹⁾.

이 중에서 logistic analysis는 Mantel과 Haenszel (1959)²⁾이 제안한 층별방법과 비교하여 볼 때, 층별의 특적인 교란인자의 영향을 조정함은 물론 한걸음 더 나아가 모델에서 교란인자로서 설정된 변수가 실제로

교란인자로 작용하고 있는지의 여부를 밝힐 수 있으며 또한 어느 인자가 효과조절인자의 역할을 할수 있는지를 가려낼 수 있는 장점을 지니고 있다. 다시 말해서 logistic analysis는 어떤 변수가 어느 정도 교란 또는 효과조절의 영향을 미치는지를 odds ratio의 값으로 수량화하여 나타내 줌으로써 질병에 대한 이해를 넓힐 수 있는 장점을 지니고 있다(Vitaliano, 1978)³⁾.

한편 log-linear analysis는 logit analysis와 유사하나 결과변수(outcome v.)인 종속변수가 독립변수들과 동일한 수준으로 처리되고, 범주변수(categorical v.)만이 분석에 사용됨으로써, 수량변수(numerical v.)를 가진 자료를 log-linear analysis에 사용하고자 할 경우 이를 범주변수로 변환시킨 후에야 분석이 가능하게 된다.

따라서 범주적인 성격을 띤 어떤 변수가 결과변수로 결정되어 있고, 교란변수에 수량변수와 범주변수가 섞여있는 형태를 가지는 역학조사자료를 분석함에 있어서 교란요인의 영향을 제거하고자 할 때에는 logistic analysis를 가장 적합한 분석기법으로 생각할 수 있다.

본 논문은 이원철(1982)⁴⁾이 속발성불임과 인공유산의 관련성을 검토하기 위하여 행한 사례-비교군 관찰 연구의 자료를 이용하여, 자연유산의 경험여부와 대상 여성의 나이를 교란요인으로 하였을 때 인공유산이 속발성불임에 미치는 영향을 밝힐 수 있는 logistic 모델을 설정하고자 시도하였다.

II. 대상 및 방법

A. 조사대상

사례군으로는 1979년 1월부터 1981년 12월까지 명동성모병원과 강남성모병원에 1년 이상의 속발성불임증을 주소로 입원 또는 내원하였던 194명의 환자를 대상으로 하였고, 대조군으로는 같은 기간내에 동일병원에 입원하였던 정상분만산모 중에서 총배임의추출법으로 선정된 247예를 대상으로 하였다.

B. 자료의 분류

본 논문에서는 총배에 의한 분석과 logistic analysis를 동시에 시도하였기 때문에 편의상 모든 변수들을 2분변수(dichotomous v.)로 다루었다. 즉 원인변수(risk v.)인 인공유산의 경험여부를 인공유산을 한번도 경험하지 않은 군(NO)과 한번 이상 경험한 군(YES)으로 나누었고 교란변수인 자연유산의 경험여부도 역시 NO, YES로 2분하였으며 또 하나의 교란변수인 대상여성의 나이는 28세 이하와 29세 이상의 2군으로 분류하였다.

분류수준을 너무 많게 잡을 경우 도수분포표에서 빈도가 0이 되는 칸(empty cell)이 많아지고 이로 인하여 분석결과의 신빙도가 떨어진다는 점(SAS Institute, 1979)⁵⁾을 감안하여 더 이상의 분류를 하지 않았다.

이원철(1982)⁴⁾의 속발성불임 자료를 우선 일원분류표로서 속발성불임군과 대조군에 대한 반응빈도 및 반응율을 나타낸 것이 표 1이다. 표에서 보는 바와 같이 점차 아래로 내려갈수록 속발성불임의 반응율이 높아져 인공유산 또는 자연유산을 경험하지 않은 군보다는 경험한 군이, 나이가 적은 군보다는 많은 군이 속발성불임이 될 가능성이 큼을 알 수 있다.

Table 1. Oneway table of secondary infertility cases and controls with two covariates

Induced abortion	Spontaneous abortion	Age	Response frequencies		Response probabilities	
			case	control	case	control
NO	NO	≤28	12	56	0.18	0.82
		≥29	23	57	0.29	0.71
	YES	≤28	19	20	0.49	0.51
		≥29	18	8	0.69	0.31
YES	NO	≤28	59	64	0.48	0.52
		≥29	39	30	0.57	0.43
	YES	≤28	9	3	0.75	0.25
		≥29	15	9	0.63	0.37

NO: never experienced

YES: experienced once or more

C. 총배방법(Mantel-Haenszel method)

이 방법은 조사하고자 하는 결과변수(outcome v.)와 원인변수(risk v.) 양쪽 모두에 관련이 있다고 생각되는 변수 즉 교란변수들을 기준으로 층을 나눈 후 각층에 해당하는 결과변수와 원인변수의 빈도를 산정한 후에 각층에서 계산된 cross product ratio (odds ratio)에 그 층의 빈도의 합계를 가중시킨 평균적인 cross product ratio를 산출함으로써 교란요인의 영향을 조정할 overall odds ratio를 구하는 방법이다(Mantel Haenszel, 1959)²⁾.

이 방법을 이용하여 대상여성의 나이와 자연유산의 경험여부가 주는 영향들을 조정할 후의 인공유산과 속발성불임의 관련성의 정도를 산출한 overall odds ratio는 2.9이었다(표 2).

표 1에 나타난 빈도를 보기 쉽도록 도식화한 것이 도표 1이다. 도표의 좌측편 즉 자연유산의 경험이 있는 두 군에 Mantel-Haenszel 방법을 적용하게 되면 이것은 자연유산을 경험한 군에서 나타나는 인공유산과 속발성불임의 관련성의 정도를, 나이의 영향을 조정할

AGE	SPONT (YES)		SPONT (NO)			
	CASE	CONTROL	CASE	CONTROL		
(<28)	(YES)	15	9	(YES)	39	30
	(NO)	18	8	(NO)	23	57
(>=29)	(YES)	9	3	(YES)	59	64
	(NO)	19	20	(NO)	12	56

Fig. 1. Stratified breakdown of those data in Table 1.

Table 2. Odds ratio for interpreting effect modification and confounding by Mantel-Haenszel method

		Spontaneous abortion		Age	
		YES	NO	≤28	≥29
Effect Modification		1.36	3.72	2.20	4.05
	Overall odds ratio	2.93			
Confounding	Secondary infertility	2.35		1.28	
	Induced abortion	0.45		0.72	

후의 overall odds ratio로 표시하는 것으로서 1.36*을 나타내었다(Fleiss, 1973)⁶⁾.

같은 방법으로 자연유산을 경험하지 않은 우측 두 군에 대하여 overall odds ratio를 낸 결과 3.72를 보였다.

만약 1.36과 3.72의 차이가 통계적으로 유의하다면 자연유산 경험하였는지의 여부에 따라 인공유산과 속발성불임의 연관 정도가 영향을 받는다는 것이고 이것은 곧 자연유산이 효과조정변수임을 뜻한다(Vitaliano, 1978)³⁾. 그러나 Mantel-Haenszel 방법에서 구한 overall odds ratio는 표준오차를 추정하기 어려운 결정을 지니고 있기 때문에(Fleiss, 1973)⁶⁾, Vitaliano (1978)³⁾의 자료에서와 같이 검정이 필요치 않을 정도로 큰 차이를 보이는 결과외에는 효과조정변수 여부의 판단이 불가능한 것으로 보인다.

또 한가지 적용이 가능한 것은(Vitaliano, 1978)³⁾도 표 1의 4군에서 각각 인공유산의 경험여부를 합하여 인공유산에 관한 변수를 없앤 후, 나이의 영향을 조정 한 후의 속발성불임과 자연유산간의 overall odds ratio를 구하면 2.35가 되고(summary $\chi^2=12.98, p<0.01$)

같은 방식으로 속발성불임의 변수를 없앤 후 나이의 영향을 조정 한 후의 인공유산과 자연유산간의 overall odds ratio를 구하면 0.45로 나타나(summary $\chi^2=12.14, p<0.01$) 자연유산은 결과변수인 속발성불임과 원인변수인 인공유산 두 변수 모두와 연관을 가지고 있음을 보여 자연유산이 교란요인임을 나타낸 반면 대상여성의 나이는 양측 모두에 유의한 연관을 나타내지 않아 교란요인으로 작용하고 있지않는 결과를 보였다(표 3).

D. logistic analysis

다원분류표(multiple contingency table)의 분석을 위하여 1935년경부터 쓰이기 시작한 logit analysis는(Fleiss, 1973)⁶⁾ 서론에서 언급된 몇가지 장점외에도 많은 교란변수를 동시에 처리할 수 있고, 모델을 통하여 추정된 통계치를 지수화(exponentiate)하여 산출된 값 자체가 결과변수와 원인변수의 관련 정도를 알려주는 odds ratio 라는 점 등의 장점을 지니고 있다(Anderson et al, 1980)¹⁾.

이 방법은 1968년에 이르러 Lewis에 의해 computer program의 사용이 가능해졌으며(Fleiss, 1973)⁶⁾ 본 논문은 SAS의 FUNCAT Procedure (SAS Institute, 1979)⁵⁾

$$* \frac{\frac{15 \times 8}{50} + \frac{9 \times 20}{51}}{\frac{9 \times 18}{50} + \frac{3 \times 19}{51}} = 1.36$$

를 이용하였다.

1. 배경

logit analysis의 배경과 방법 등은 Bishop들(1975)²⁾, Anderson들 (1980)¹⁾, Breslow (1976)³⁾, Prentice (1976)⁴⁾, Breslow과 Powers (1978)¹⁰⁾ 등에서 찾아 볼 수 있다.

logit analysis는 일차회귀분석 및 공분산분석과 유사한 점을 지니고 있다(Anderson, 1980)¹⁾.

일반적으로 기본적인 공분산분석모델은 각 군에서 결과변수의 평균이 교란변수의 일차함수로서 표시된다는 가정이다. 즉 대조군과 처리군에서 각각

$$E(Y) = \alpha_0 + \beta X; \text{대조군}$$

$$E(Y) = \alpha_1 + \beta X; \text{처리군} \dots \dots \dots (1)$$

로 표현된다.

위의 두 식을 통합해서 한 식으로 나타내기 위하여 원인변수가 나타나는 경우(처리군)를 $R=1$, 나타나지 않는 경우(대조군)를 $R=0$ 로 하면 (1)식은 다음과 같이 된다.

$$E(Y) = \alpha_0 + \gamma R + \beta X \dots \dots \dots (2)$$

(여기서 $\gamma = \alpha_1 - \alpha_0$)

결과변수가 2분적인 경우에 (2)식을 그대로 적용하려고 할 때, 결과변수의 평균인 좌변은 그값이 비율이므로 분포는 0과 1 사이에 한정되어야 하는데 우변은 교란변수 X 의 값에 따라서는 얼마든지 커지거나 작아질 수도 있다는 난점이 있다.

이러한 난점을 극복하기 위하여 원인변수를 R , 교란변수를 X 로 할때 어떤 결과가 나타나는 확률 즉 Y 가 1이 되는 확률을 $P(R, X)$ 로 표시하고 이 $P(R, X)$ 의 log odds 즉 $P(R, X)$ 의 logit transformation을 결과변수의 함수로 취하면 (2)식 대신에

$$\ln \left[\frac{P(R, X)}{1 - P(R, X)} \right] = \alpha + \gamma R + \beta X \dots \dots \dots (3)$$

로 되어 logistic regression의 형태를 갖추게 된다.

이때에 처리군 ($R=1$)과 대조군 ($R=0$)사이의 odds ratio ω 는 (3)식의 Parameter γ 와 다음과 같은 관계를 가진다. 즉

$$\omega = \frac{P(1, X)}{1 - P(1, X)} \bigg/ \frac{P(0, X)}{1 - P(0, X)}$$

$$= \frac{\exp(\alpha + \gamma + \beta X)}{\exp(\alpha + \beta X)} = \exp(\gamma) \dots \dots \dots (4)$$

따라서 logistic analysis로써 γ 의 추정치 $\hat{\gamma}$ 을 알아내면 $\hat{\omega} = \exp(\hat{\gamma})$ 으로서 교란인자의 영향을 조정한 후의 원인변수와 결과변수와의 연관성의 척도인 odds ratio를 추정할 수 있다(Anderson et al., 1980)¹⁾

2. 적합한 모델선정

적합한 모델을 선정한다는 것은 회귀 분석의 경우와 마찬가지로 대상연구자료를 가장 잘 설명하는 식을 구하는 것으로, logistic analysis의 경우에는 위에 언급한 (3)식의 우변이 어떠한 변수들로 구성되어야 가장 적합한가를 정하는 과정이다(표 3).

적합한 모델을 선정하려면 첫째, 임의로 선정된 모델에 의해 추정된 모수추정치들 가지고 예측빈도를 계산하고 둘째, 예측빈도와 실측빈도로 산출한 χ^2 의 값을 근거로 하여 선정된 모델이 적합한지를 평가한다. 실제적으로는 위 두 단계는 computer program에 의하여 수행되어지므로 이 과정을 여러가지 가능한 모델에 적용함으로써 가장 만족할만한 모델을 찾아내게 된다(Anderson et al. 1980)¹⁾.

모델에 포함시키고자 하는 어떤 특정한 변수의 효과가 유의한지를 검정하기 위하여 표 3에서 볼 수 있는 바와 같은 방법을 적용할 수 있다.

즉 본 자료에서 고려하고 있는 변수의 주효과만을 모델에 포함시킨 1 번모델의 χ^2 에서, 여기에 인공유산의 경험여부와 자연유산의 경험여부의 교차효과변수를

Table 3. The alternative models among which the model for secondary infertility with induced abortion is chosen

Model	Model Fitting			Testing a variable		
	DF	χ^2	P	effect tested	difference in χ^2 for models	significance
1. (IN)(SP)(AG)	4	6.69	0.15	—	—	—
2. (IN)(SP)(AG)(IN*SP)	3	2.66	0.44	(IN*SP)	4.03†	$P < 0.05$
3. (IN)(SP)(AG)(IN*SP)(IN*AG)	2	1.33	0.51	(IN*AG)	1.33	NS
4. (IN)(SP)(AG)(IN*SP)(IN*AG)(SP*AG)	1	1.16	0.28	(SP*AG)	0.17	NS
5. (IN)(SP)(AG)(IN*SP)(IN*AG)(SP*AG)(IN*SP*AG)	0	0.00	1.00	(IN*SP*AG)	1.16	NS

(IN): main effect of Induced abortion

(SP): main effect of Spontaneous abortion

(AG): main effect of Age

(IN*SP): interaction effect between Induced abortion and Spontaneous abortion

†Each of these tests contains one degree of freedom

Table 4. Estimation and testing of the parameters by a logistic model

Paramete	DF	Estimate	χ^2	P	STD† of Estimate
Intercept	1	0.047	0.14	0.706	0.125
Induced abortion (IN)	1	0.392	9.76	0.002	0.125
Spontaneous abortion (SP)	1	0.501	15.91	0.0001	0.125
Age (AG)	1	0.219	4.28	0.039	0.106
(IN*SP)††	1	0.260	4.03	0.045	0.130
Residual	3		2.66	0.447	

† standard deviation

†† Interaction between Induced abortion and Spontaneous abortion

Table 5. Observed frequencies and expected frequencies of the chosen model

Spontaneous abortion		NO		YES		
Age		≤28	≥29	≤28	≥29	
Induced abortion	NO	Observed disease	12.00	23.00	19.00	18.00
		Expected disease	13.24	21.76	20.55	16.45
		Total observed	68.00	80.00	39.00	26.00
		Estimated relative risk	1.00	1.55	4.61	7.12
Induced abortion	YES	Observed disease	59.00	39.00	9.00	15.00
		Expected disease	58.00	40.00	7.21	16.79
		Total observed	123.00	69.00	12.00	24.00
		Estimated relative risk	3.69	5.70	6.23	9.63

점까지 2번 모델의 χ^2 를 빼면 첨가된 교차효과변수에 의하여 설명되어지는 χ^2 를 구할 수 있어 자유도 1로써 이를 검정할 수 있게 된다.

이러한 과정을 통하여, 좋은 적합도를 유지하면서 가장 간단한 변수들로 구성된 모델을 선정하는 것이 이 단계의 목적인 바(Fienberg, 1977)¹¹⁾ 본 자료에서는 표 3의 2번 모델이 변수 모두가 유의함을 나타내고 모델의 적합도 역시 $\chi^2=2.66$ $P=0.44$ 로 좋은 결과를 나타내어 본 자료를 가장 잘 설명하여 주는 모델로 선정하였다.

표 5는 이 모델로 추정된 모수추정치를 이용하여 예측빈도와 실측빈도를 비교한 표로써, 실측빈도는 모두 예측빈도에 가까운 값을 나타내어 모델이 자료에 적합함을 보여주는 증거가 된다. 즉 예측빈도와 실측빈도로 산출한 χ^2 의 값이 클수록 또 이에 해당하는 P 값이 적으면 적을수록 모델의 적합도는 떨어지는 것이다. 예측빈도와 실측빈도를 가지고 모델의 적합성을 판정하는 방법에는 Person의 χ^2 statistic과 likelihood ratio statistic이 있다(Anderson et al, 1980)¹²⁾.

3. 교란변수(confounding variable)와 효과조정변수(effect modifying variable)

logistic 모델내에서의 보조변량(corariates)의 주효과(main effect)들에 대한 분석은(본 자료에서는 (AG)와 (SP)가 해당됨) 질병의 유무 등으로 표현되는 결과변

수와 각 보조변량 사이의 연관성의 정도를 나타내 주기 때문에 각 보조변량의 주효과가 모델에서 유의하게 검정되었다는 것은 곧 그 보조변량이 교란요인의 역할을 한다는 것을 의미한다(Vitaliano, 1978)³⁾.

질병과 보조변량 사이의 연관성의 정도를 나타낸다는 의미는 표 4에서 쉽게 이해될 수 있다. 표 4는 앞에서 가장 적합하다고 선정된 모델의 각 변수에 해당하는 모수를 추정하고 검정한 FUNCAT Procedure의 결과를 옮긴 표이다.

앞에 언급한 (4)식에서 r 를 산출해 내는 방식을 이용하여 원인변수의 추정치 0.392와 원인변수가 포함된 교차효과(interaction effect)의 추정치 0.260을 합한 0.652를 지수화(exponentiate) 했을 때 산출되는 값 1.92는 교란요인의 영향을 조정했을 때의 인공유산과 속발성불임의 관련정도를 나타내주는 odds ratio 이다(Anderson et al, 1980)¹²⁾.

즉 추정치의 값이 크다는 것은 그만큼 결과변수에 주는 영향이 크다는 것을 의미하게 되므로 표에서 보는 바와 같이 자연유산이 대상 여성의 나이보다는 속발성불임에 더 큰 영향을 주고 있음을 파악할 수 있다.

지금까지 언급한 교란요인의 문제는 원인변수의 영향이 항상 일정하게 결과변수에 나타난다는 가정하에(no-interation, parallelism) 성립되었으나 실제적으로

Table 6. Odds ratio calculated from the estimates of the logistic analysis

Spontaneous Abortion	NO		YES	
	≤28	≥29	≤28	≥29
Age				
Induced Abortion	NO	1.00	4.57	7.10
	YES	3.67	5.70	5.93

는 많은 경우에 있어서 원인변수의 영향이 어떠한 배경요인(background factor)에 의해 좌우된다. 이것은 원인 변수가 나타날 경우와 나타나지 않을 경우의 두 직선의 기울기가 다른 경우를 의미하게 되며 이런 경우 원인변수는 배경요인과 함수관계를 가지게 되고 logistic 모델에서는 (IN* SP)의 교차효차와 같은 형태로 나타나게 된다.

이때 원인변수에 영향을 주는 배경요인을 곧 효과조정변수로 볼수 있는 바 자연유산의 경험 여부가 이어해당된다. 이 모델에서와 같이 모델에서의 교란요인이 동시에 효과조정 변수로도 작용하는 경우가 있고 모델에서는 교란요인으로 작용하지 않으나 효과조정요인으로 작용하는 경우도 있다(Anderson et al, 1980; Vitaliano, 1978)^{11,12}.

교란변수의 경우와 마찬가지로 한 모델에서 효과조정변수가 2개 이상 있을 경우 어느 변수가 더 크게 작용하고 있는지는 교차효차의 추정치에 의해 비교될 수 있다.

4. 추정 및 검증

모수를 추정하는 방법에는 maximum likelihood estimation, minimum χ^2 estimation, minimum logit χ^2 estimation, multivariate discriminant analysis 등이 있는 바 SAS의 FUNCAT은 minimum logit χ^2 estimation 방법을 사용하고 있다.

표 4는 앞에서 선정된 모델에 있어서의 모수추정과 검정을 computer program이 행한 결과를 옮긴 표이다.

이러한 모수추정치로 산정된 예측빈도와 실측빈도의 값을 이용하여 산출된 χ^2 가 표 4의 Residual으로써 모델의 적합성을 평가하는 기준이 된다는 것은 이미 언급하였다.

Intercept로 표현된 추정치는 일차회귀직선에 있어서의 절편과 같은 것이나 변수들이 어떠한 형태를 취하는가에 따라 값이 변하기 때문에 일반적인 해석이 불가능한 것으로 알려져 있다(Anderson, 1980)¹¹.

III. 결 과

표 4에서 얻어진 결과를 가지고 (4)식에서 r 를 산출해내는 방식을 이용하여, 인공유산과 자연유산을 한번도 경험한 적이 없고 나이가 28세 이하인 군을 기준으

로 하였을 때 각 군이 나타내는 odds ratio를 산출한 결과가 표 6이다.

표 1에서 나타났던 바와 같이 인공유산과 자연유산을 경험한 군이 경험하지 않은 군보다, 나이가 많은 군이 적은 군보다 속발성불임이 될 가능성이 높았던 경향이 이 표에서는 odds ratio를 통하여 표현되었다.

앞에서 언급되어진 바와 같이 나이가 주는 영향보다는 자연유산이 주는 영향이 더 큼을 표에서 볼 수 있으며 자연유산을 경험하지 않은 군에서 인공유산의 경험여부에 따른 odds ratio의 차이가, 자연유산을 경험한 군에서 인공유산의 경험여부에 의해 나타나는 odds ratio의 차보다 더 큰 비율을 나타내 자연유산이 효과조정요인으로 작용함을 나타내었다.

전반적으로 인공유산보다는 자연유산이 속발성불임과 더 큰 연관을 보이고 있는 것은 자연유산의 반정도가 난자의 이상으로 생길 수 있어 (Hertig & Sheldon, 1943)¹² 외적조건에 의하여 영향받는 인공유산보다는 위와 같은 신체적인 결합으로 인하여 야기될 수 있는 자연유산을 경험한 여성이 속발성불임이 될 가능성이 높으리라는 설명과 일치한다(이원철, 1982).

IV. 고 찰

Vitaliano (1978)²²는 피부암의 발생과 태양에 폭로된 시간과의 관련성을 밝히기 위하여 행한 논문에서 Mantel-Haenszel 방법과 Logistic regression analysis를 이용하여 어떤 변수가 교란요인인지, 효과조정요인인지를 밝히는 것을 시도하였다. Vitaliano (1978)²³의 결과는 두가지 방법에서 모두 동일한 양상을 띄었으나 본 논문에서는 Mantel-Haenszel 방법의 적용에 문제가 있음을 지적하였을 뿐 동일한 결과를 얻지 못하였다.

실제 문제의 적용에 있어서도 대상여성의 나이는 인공유산과 속발성불임에 관련해서 교란인자로 작용할 수 있음이 알려져 있고 (WHO, 1978)¹³ Trichopoulos (1976)¹⁴들의 대상여성의 나이와 사회경제적 지위 그리고 임신경험회수를 고려하여 환자군과 대조군을 짝지은 후 인공유산 및 자연유산의 경험여부를 동시에 비교하여 행한 조사에서도 자연유산의 경험여부에 따라 비교위험도가 크게 변함을 나타내어 자연유산이 효과조정요인으로 작용할 수 있음을 볼 수 있었다.

Logistic analysis를 행할 수 있는 Computer program 중의 하나인 Ronald들¹⁵⁾에 의한 LINCAT program은 일정한 자료에 대하여 linear regression, log-linear regression, logistic regression을 동시에 처리할 수 있는 장점을 가지고 있기는 하나 실제 자료를 적용시킬 경우에는 일단 빈도표(frequency table)를 만들어야 하고 복잡하지는 않다 하더라도 프로그램을 작성해야 하는 난점이 있다.

본 자료와 같이 사례-비교군 관찰연구(case-control study)로서 후향적으로 추출된 자료를 가지고 logistic 모델을 이용하여 전향적으로 분석하는 것에 대하여, Breslow와 Powers(1978)¹⁰⁾는 모든 가능한 교란 요인들을 보조변량으로 하여 모델에 도입시켰을 경우에는 후향적인 자료에서 추정된 값과 전향적인 자료에서 추정된 값이 일치하다고 하였다. 본 논문에서 교란요인으로 선정했던 자연유산의 경험여부나 대상여성의 나이 이외에도 본자료에서 다루지 못한 대상여성의 사회경제적 요소라던가 인공유산당시의 임신개월수 등의 가능한 교란요인이 될 수 있는 영향들은 감안 할 때, 앞에서 산출된 odds ratio는 단순히 관련의 정도를 나타내어 줄 뿐 그 값 자체를 비교위험도를 추정하여 해석하기에는 무리가 있을 것으로 보인다.

Anderson들(1980)¹¹⁾은 이 문제의 해결을 위하여 원인 변수와 결과변수의 역할을 바꾸어 원인변수를 logistic 모델에 있어서의 종속변수에 대입하고, 결과변수를 독립변수들 중의 한 변수도 취급하는 방법을 제안하였다.

Mantel(1973)¹⁶⁾은 자료가 사례-비교군 방법에 의하여 추출되었다 하더라도 모집단을 대표할만큼 확률적으로 추출된 경우에는 logistic model의 사용에 무리가 없다고 하였다.

V. 결 론

속발성불임과 인공유산에 관한 사례-비교군 관찰연구의 자료를 이용하여 logistic analysis를 적용함으로써, 자연유산의 경험 여부와 대상여성의 나이를 교란요인으로 하였을 때 인공유산이 속발성불임에 미치는 영향을 밝힐 수 있는 logistic 모델을 설정하고자 본 연구를 시도하였다.

SAS의 FUNCAT Procedure를 이용하여 분석한 결과 인공유산의 경험여부, 자연유산의 경험여부, 대상여성의 나이, 그리고 인공유산과 자연유산 경험여부의 교차효과변수가 독립변수들로 구성되는 logistic 모델이 본 자료를 가장 잘 설명하여 주는 모델로 선정되었다.

logistic analysis는 교란인자의 영향을 조정함은 물론, 실제로 그 변수가 교란인자로 작용하고 있는지의 여부를 밝힐 수 있으며 또한 어떤 인자가 효과조정인자의

역할을 할 수 있는지를 가려내는 장점을 지니고 있는바 본 자료에서는 자연유산의 경험여부가 교란인자와 효과조정인자의 역할을 할 수 있음을 보였으며 대상여성의 나이는 교란 인자로 작용할 수 있음을 나타내었다.

(본 논문의 작성에 있어서 큰 도움을 주신 이흥준교수님과 송혜향교수님께 깊은 감사를 드립니다).

—References—

1. Anderson, H., Auquier, A., Hauck, W.W., Oakes, D., Vandaele, W. & Weisberg, H.I.: *Statistical Methods for Comparative Studies*, John Wiley & Sons, New York, 1980.
2. Mantel, N. & Haenszel, W.: *Statistical aspects of the analysis of data from retrospective studies of disease*, *J. Nat. Cancer Inst.*, 22 : 719~748, 1959.
3. Vitaliano, P.P.: *The use of logistic regression for modelling risk factors: With application to non-melanoma skin cancer*, *Am. J. Epidemiol.*, 108 : 402~414, 1978.
4. 이원철 : 인공유산 및 자연유산이 속발성불임에 미치는 영향에 관한 사례-비교군 관찰연구, 가톨릭大學 醫學部 論文集, 35 : 557~563, 1982.
5. SAS Institute: *SAS User's guide 1979 Edition*, SAS Institute Inc., North Carolina, 1979.
6. Fleiss, J.L.: *Statistical Methods for Rates and Proportions*, John Wiley & Sons, New York, 1973.
7. Bishop, Yvonne M.M., Fienberg, S.E. and Holland, P.W.: *Discrete Multivariate Analysis: Theory and Practice*, London, 1975.
8. Breslow, N.: *Regression analysis of the log odds' ratio: A method for retrospective studies*, *Biometrics*, 32 : 409~416, 1976.
9. Prentice, R.: *Use of logistic model in retrospective studies*, *Biometrics*, 32 : 599~606, 1976.
10. Breslow, N. & Powers, W.: *Are there two logistic regressions for retrospective studies?*, *Biometrics*, 34 : 100~105, 1978.
11. Fienberg, S.E.: *The Analysis of Cross-Classified Categorical Data*, 2nd Ed., The MIT Press, Cambridge, 1977.
12. Hertig, A.T. & Sheldon, W.H.: *Minimal criteria required to prove prima facie case of traumatic*

- abortion or miscarriage, Ann. Surg.* 117 : 596, 1943 Cited from Pritchard, J.A. & Macdonald, P.C.: *Williams Obstetrics, 6th Ed.* Appleton-Century-Croft, New York, 1980.
13. WHO: *Induced abortion. Technical report series No. 623, Geneva, 1978.*
 14. Trichopoulos, D., Handanos, N., Danezis, J., Kalandidi, A & Kalapothaki, V.; *Induced abortion and secondary infertility. Brit. J. Obstet. Gynecol.* 83 : 645~650, 1976.
 15. Ronald, N., Forthofer, C., Starmer, F. and Grizzle, J.E.: *A Program for the Analysis of Categorical Data by Linear Models.* Houston.
 16. Mantel, N.: *Synthetic retrospective studies and related topics, Biometrics,* 29 : 479~486, 1973