

의료이용도에 대한 인구학적 변수의 효과분석의 방법

한림대학교 사회의학연구소
김병익 · 이영조 · 권순호 · 한달선

<Abstract>

Analysis of the Effects of Demographic Variables on Health Care Services Using the Spline Regression

Byungik kim, Youngjo Lee, Soonho Kwon, Dalsun Han

Institute of Health Services and Management, Hallym University

Demographic variables have a great deal of impact on the utilization of health services. In this paper, the use of segmented polinomials is shown to be superior to the simple use of dummy variables and simple polinomials in explaining differences in health care utilization with respect to sex and age differences.

Key Words : Segmented Polinomials, Simple Polinomials, Demographic Variables, Utilization.

1. 서 론

의료이용도에 영향을 주는 요인들로는 질병과 관련된 의료필요도 요인, 경제적 요인, 사회인구학적 요인, 그리고 기타 요인 등으로 대별되는 데, 그 중 인구학적 변수로는 성과 연령이 대표적이라고 볼 수 있겠다. 연령별 질병발생은 생물학적 특성에 의해 연령이 높아질수록 만성병의 발생이 많으며 반대로 연령이 낮을수록 급성상병의 발생이 많은 데, 선진국의 경우 의료이용도는 10대 후반 또는 20대 연령에서 가장 낮고 0~4세 및 70세 이상에서 가장 높은 U자형의 곡선 분포를 이룬다고 보고되고 있다(Ohmura, 1978).

현재 우리나라는 노인의 의료이용도가 다른 연령층에 비해 그리 높지 않기 때문에 U자형이 아닌 다른 곡선모양을 이룰 것으로 여겨진다. 그러나 의료보험이 전국민에게 확대됨에 따라 미충족 의료요구가 높은 노인들의 의료이용이 높아지면 선진국과 같은 U자형 곡선으로 변화하리라고 본다.

남녀별 의료이용도의 양상은 근본적으로 차이가 별로 없을 것으로 기대되나 20대 이후의 여자는 임신, 출산 및 관련 질병, 남녀 수명차 등으로 인해 약간의 차이를 나타낼 것으로 생각된다.

따라서 의료이용도를 보다 잘 설명하기 위해 성과 연령별로 보다 구체적으로 추정할 필요가 있다. 지금까지의 실증분석을 보면 성은 남과 여를 구분하는 가변수(dummy variable)로, 연령은 의료이용도와 직선관계를 가정하여 추정한 결과가 적지 않았다. 연령에 따른 의료 이용도를 좀 더 구체적으로 분석한 연구로서는 연령을 구간별로 가변수 처리하거나 2차 또는 3차 함수로서 추정한 것이 있을 뿐이었다(이헌협, 1982 ; Lee, 1984 ; 노공균, 1986 ; 한달선의, 1986 ; Van der Gaag, 1978).

본연구는 성, 연령별 의료이용도의 분포를 그림으로 나타내보고(plotting), 여기에 가장 적합한 함수를 도출해 냄으로써 성과 연령에 의한 의료이용의 설명도를 최대한 높이기 위한 시도를 해본 것이다. 또한 다른 연구에서 측정한 성과 연령 변수에 비해 얼마 만큼의 설명도 차이를 보이는지도 검토해 보았다. 앞으로 우리나라도 인구의 노령화가 가속화되어 인구구조가 바뀌고 있음은 감안할 때, 이러한 방법을 통해 인구학적 변수들이 의료수요에 미치는 영향을 보다 정확히 이해할 필요성이 커지고 있는 것이다.

2. 모형설정을 위한 탐색

생존곡선분석(Survival Analysis)에서 연령별 생존위험도 곡선(Hazard Function)은 U자형 또는 욕조형(Bathtub-Shape)을 나타낸다(Lawless, 1983). 질병발생에 대한 위험도 곡선도 욕조형을 가질 것으로 생각되므로, 연령에 따른 의료이용도도 U자형의 곡선형태를 가질 것으로 생각된다. 그러나 의료이용도가 단지 위험도에 의해서만 결정되는 것이 아니므로 여러가지 요인들이 작용하였을 때 연령별 의료이용도는 다른 모양의 곡선관계를 보일 수도 있다.

성별, 연령별 의료이용도는 어떤 함수로 적절하게 설명될 수 있는지 살펴보고, 성과 연령 및 다른 변수들에 의해 의료이용도를 설명하기 위해 <표 1>과 같이 서울시내 3개 직장의료보험 조합의 1985년 1년간 연속자격보유자의 5% 표본자료를 각각 사용하였다.

<표 1> 자료와 분석대상 (단위 ; 명)

조합명	85년말 적용인구	연간 연속 자격보유자(%) ¹⁾	Logit분석대상자 (%) ²⁾
수출공단 서울지구	120,767	76,407(63.3)	3,810(5.0)
서울 제 3 지구	145,288	86,752(59.7)	4,327(5.0)
서울 제 8 지구	163,475	115,864(70.9)	5,779(5.0)
총 계	429,530	279,023(65.0)	13,916(5.0)

주 : 1) 연말 적용인구에 대한 비율
2) 연속 자격보유자에 대한 비율

개인의 의료이용 경험여부를 종속변수로 하여 성, 연령 등의 인구학적 변수, 보수월액 등의 경제적

요인, 피보험자의 성과 연령, 식구수, 가족내 성인의 비율, 보험가입기간 등을 독립변수로 선정하여 분석하려고 한다. 이 경우 각 개인의 의료이용 경험 여부는 '이용' 및 '비이용'의 두가지 값만 갖게 되므로 종속변수가 이항분포를 따른다고 볼 수 있다. 그러므로 McCullagh와 Nelder(1989)에 따라 적당한 연관함수(link function)를 설정하여서 모형을 확정하였다.

연관함수로써는 이런 경우 흔히 사용되는 Logit를 이용하였다. 각 개인의 의료이용 확률을 P라 하고 해당하는 독립변수 들을 $X_1, X_2, X_3, \dots, X_n$ 이라고 하자. 이 경우 모형은

$$\log \frac{P}{1-P} = f(X_1, X_2, X_3, \dots, X_n)$$

이다. 보다 좋은 모형을 설정하기 위하여는 종속변수 즉, 의료이용 여부와 독립변수 상호간의 연관 관계에 대한 보다 면밀한 검토가 요구된다.

이 논문에서는 인구학적 변수에 따른 의료 이용도의 변화양태에 초점을 맞추었다. 이를 위해 분석대상자 279,023명 전원을 여자(151,467명)와 남자(127,556명)로 구분하고 각각에 대해 연령별로 의료이용 확률 P^* 를 구한 후 P^* 의 Logit 즉, $\log \frac{P^*}{1-P^*}$ 를 매 연령별로 아래 그림 1과 그림 2와 같이 그려 보았다.

그림에서 관측치들은 *로 표시되어 있는 데, 여자의 경우 의료이용이 십대 후반까지 급격히 감소하다가 20대부터 30대 초반까지 임신, 출산 등에 의해 상승하며 그후 40대후반까지 감소하다가 폐경기를 전후한 50대 초에 다시 증가한 후 곧 감소하는 것을 알 수 있다(그림 1).

남자의 경우는 십대 후반까지 급격히 감소하고 20대 부터 30대 까지 다시 증가하다가 30대 이후에는 그 증가추세가 둔화, 정체되는 것 같다(그림 2).

본 자료에 의하면 현재 한국의 의료이용률은 노령화됨에 따라 의료이용도가 증가한다는 Ohmura(1978)의 U자형 곡선과는 다른 것을 알 수 있다. 오히려 노년층의 의료이용도는 연령에 따라 증가하는 것이 아니라 감소(여) 또는 정체(남)하는 경향을 나타낸다. 그러나 전국민 의료보험 등에 의해 의료이용이 용이해지면서 노인의 의료이용이 많이 증가하면 점차 U자형에 접근할 것으로 생각된다.

그림에서 보는 바와 같이 곡선형태는 어떤 특정한 함수를 이용해 표현하기는 불가능하다. 따라서 이 경우 접목된 다항식(Segmented Polynomial)을 사용하는 것이 바람직하다(Prenter, 1975 ; Sard and Weintraub, 1971).

- 1) 남자의 경우는 35세 이후 뚜렷한 양상(Pattern)이 없이 의료 이용도가 다소 변동하는데, 본 논문에서는 일차식을 이용하여 설명하였다. 35세 이후의 구간을 좀더 세분화하여 2차 또는 3차 곡선으로 추정해 볼 수도 있겠다. 이 경우에도 마지막 구간의 함수는 1차 함수가 선정되는 것이 바람직하다. 그 이유는 74세 이후의 의료이용율을 예측하고자 할 때(extrapolation), 2차 또는 3차 함수가 선정되면 예측치가 급격히 변하여 실제와 매우 다를 수 있기 때문이다.

본 논문에서는 여자의 경우 25세, 35세, 55세를 접목점으로 하여 연령을 4구간으로 나누고, 각 구간에서 2차함수를 이용하여 의료이용도를 근사하였다. 단, 함수를 매끄러운 연속 함수로 하기 위해, 각 접목점에서 함수들이 같은 값을 가지며 그 미분값이 같도록 하여, 그 함수의 계수를 회귀분석으로 구하면 그림 1과 그림 2에서와 같이 'P'로 나타난다. <표 2>와 <표 3>에서 보자시피

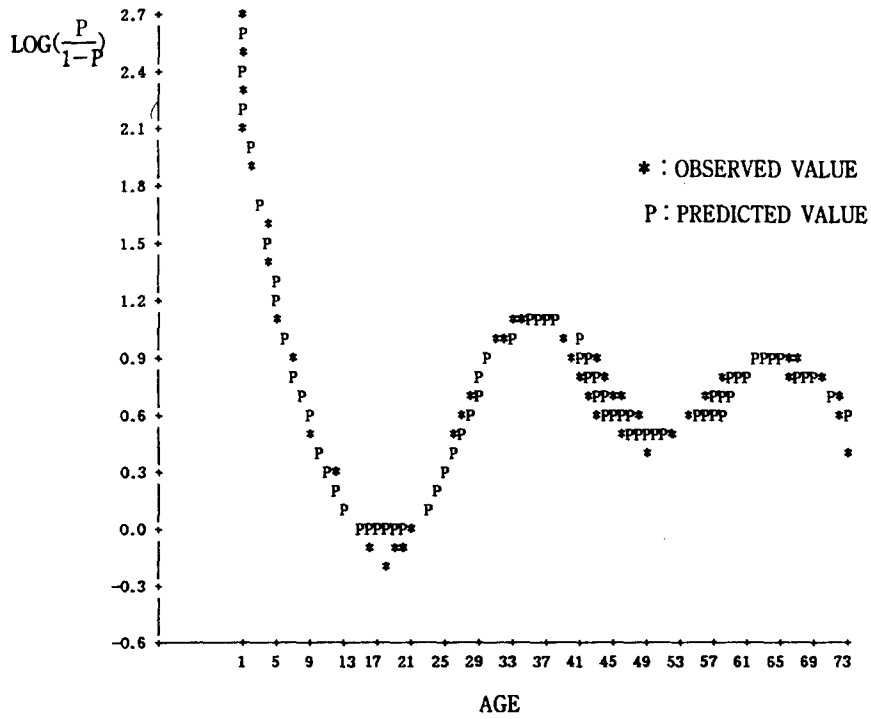


그림 1. 여자의 연령별 의료이용도(N=151,467명)

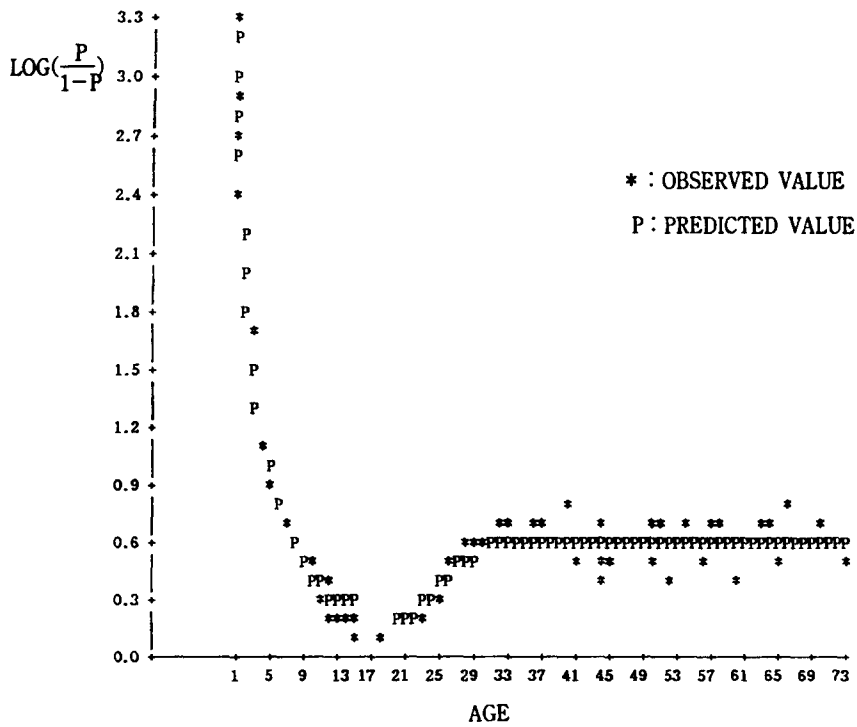


그림 2. 남자의 연령별 의료이용도(N=127,556명)

R-Square 값이 여자의 경우 98.6%, 남자의 경우 98.8%로 접목된 다항식이 의료이용 경험률을 잘 설명하는 것을 알 수 있다.

여자의 경우는 4개의 구간에서 각각 2차항수를 잡았으므로 각 구간당 계수가 3개씩이므로 총 12개의 계수가 있다. 그러나 3개의 접목점에서 함수들이 연결되고 미분값을 같게 하는 조건을 각기 2개씩 주었으므로 총 6개의 계수가 소거되어 6개의 변수만이 남는다.

〈표 2〉 여자의 경우 접목된 회귀분석

R-SQUARE=0.9857

DEPENDENT VARIABLE : $\log(P^*/1-P^*)$

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F
MODEL	6	62.0407	10.3401	793.18	0.0001
ERROR	69	0.8995	0.0130		ROOT MSE
UNCORRECTED TOTAL	75	62.9402			0.1142

PARAMETER ¹⁾	ESTIMATE	T FOR HO : PARAMETER=0	PR> T	STD ERROR OF ESTIMATE
변수1	2.9805	44.38	0.0001	0.0672
변수2	-0.3221	-32.49	0.0001	0.0099
변수3	0.0091	29.81	0.0001	0.0003
변수4	-8.5076	-18.47	0.0001	0.4606
변수5	-7.3573	-5.40	0.0001	1.3627
변수6	-0.0021	5.65	0.0001	0.00037

1) 변수1 : 절편(0-25세 구간)

변수2 : 1차항의 계수(0-25세 구간)

변수3 : 2차항의 계수(0-25세 구간)

변수4 : 절편(25-35세 구간)

변수5 : 절편(55세 이상 구간)

변수6 : 2차항의 계수(55세 이상 구간)

남자의 경우는 35세를 접목점으로 35세 이하는 3차항수, 35세 이상은 1차항수로 정하였다. 첫번째 구간에서 4개의 계수가 있고 두번째 구간에서 2개의 계수가 있으나 35세에서 상기의 두조건을 주었으므로 총 변수는 4개만 남게 된다. 이 가운데 변수4의 t값이 유의하지 않은 데 35세 이후의 의료이용도가 연령에 관계없이 거의 같음을 의미한다.

3. 접목된 다항식을 이용한 Logistic 분석

전장에서 연령에 따른 의료이용율의 Logit 값을 흔히 사용되는 함수로 표시할 수 없으므로 접목된

〈표 3〉

남자의 경우 접목된 회귀분석

R-SQUARE=0.9879

DEPENDENT VARIABLE : log (P*/1-P*)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F
MODEL	4	51.7948	12.9487	1454.91	0.0001
ERROR	71	0.6319	0.0089		ROOTMSE
UNCOPRECTED TOTAL	75	52.4267			0.0943

PARAMENTER ¹⁾	ESTIMATE	T FOR HO : PARAMETER=0	PR> T	STD ERROR OF ESTIMATE
변수1	3.2944	52.87	0.0001	0.0623
변수2	-0.3897	-33.44	0.0001	0.0117
변수3	0.0156	27.33	0.0001	0.0006
변수4	0.0004	0.44	0.6649	0.0010

- 1) 변수1 : 절편(0-35세 구간)
- 변수2 : 1차항의 계수(0-35세 구간)
- 변수3 : 2차항의 계수(0-35세 구간)
- 변수4 : 1차항의 계수(35세 이후 구간)

다항식으로 설명하는 것이 타당함을 보여주었다. 그러므로 Logistic 회귀분석을 수행할 때 성, 연령 변수는 전장에서 결정된 성별 접목된 다항식을 사용하여 측정하는 것이 바람직하겠다. 〈표 4〉는 개인의 의료이용 경험의 Logistic 회귀분석 결과를 보여주고 있는 데, 〈표 1〉에서 제시된 자료 중 5% 즉, 13,916명을 무작위로 추출한 표본을 대상으로 분석한 것이다.

전체 모형의 절편이 구해졌으므로 접목된 다항식의 변수가 하나 줄어든 것을 유의하여야 한다. 이 분석에서의 절편값은 〈표 2〉의 여자인 경우 접목된 회귀분석의 변수1 즉, 첫번째 구간 2차함수의 절편이 된다. 그러므로 남자의 접목된 다항식에서 변수1은 남자와 여자의 절편값의 차이를 나타내며 이 추정치가 유의한 양의 값을 가지므로 남자가 여자 보다 출생시 의료이용률이 높음을 나타낸다.

총 23개의 독립변수를 포함했을 때 Chi-Square 값은 1062.22였다. 유의하지 않은 독립변수인 피보험자의 연령, 피보험자와의 가족관계, 남자 35세 이상의 경우 연령증가에 따른 의료이용도의 증감을 나타내는 변수4 등을 제외한 최종모형을 사용하여도 Chi-Square 값이 1056.56으로 여전히 높았다.

동일한 자료를 이용하여 성, 연령 변수를 종래의 방법대로 가변수 처리, 1차 및 3차함수로 잡은 경우와 본연구의 결과를 비교한 것이 〈표 5〉에 제시되어 있다. 접목된 다항식을 사용한 식이 의료이용도를 설명하는 데 가장 적합한 것으로 나타났다. 특히, 연령을 단순히 1차함수로 처리한 경우와 접목된 다항식을 이용한 최종모형의 Chi-Square 값을 비교해보면 차이가 매우 크므로 의료이용도를

<표 4>

년간 의료이용경험 여부

VARIABLE		BETA	STD. ERROR	CHI-SQUARE	P
절편		2.4524	0.2301	113.59	.
	0-29	0.1262	0.0691	3.33	0.0680
피보험자 연령	40-29	0.0674	0.0648	1.08	0.2988
	50+	0.1895	0.0981	3.73	0.0535
피보험자 성관계	남자	0.1998	0.0606	10.85	0.0010
	배우자	0.0940	0.0860	1.19	0.2745
	존속	0.1293	0.1354	0.91	0.3397
	비속	-0.2485	0.1315	3.57	0.0587
보수월액		0.0072	0.0009	59.94	0.0000
가족수		-0.0822	0.0152	29.13	0.0000
성인가족비율		-0.5045	0.0747	45.56	0.0000
	1-3년	0.2884	0.0471	37.39	0.0000
가입기간	3-5년	0.4338	0.0581	55.56	0.0000
	5-7년	0.8060	0.0782	105.98	.
	7년 초과	0.7236	0.0747	93.81	.
	변수1	0.6788	0.2331	8.48	0.0036
남자의	변수2	-0.3787	0.0346	119.76	.
접목된 다항식	변수3	0.0148	0.0017	75.60	.
	변수4	-0.0081	0.0051	2.51	0.1133
	변수2	-0.2439	0.0217	126.28	.
여자의	변수3	0.0064	0.0006	97.97	.
접목된 다항식	변수4	-7.4145	1.0880	46.43	0.0000
	변수5	-5.8299	2.0687	7.94	0.0048
	변수6	-0.0011	0.0005	4.77	0.0290

CHI-SQUARE=1062.22 WITH 23 DF

기준변수 : 피보험자 연령 : 30-39, 성 : 여자, 가입기간 : 0-1년

<표 5>

성, 연령변수 처리방법별 LOGISTIC REGRESSION 결과 비교

처리방법	CHI-SQUARE	DF
가변수처리 ¹⁾	965.67	20
1차함수	764.17	16
3차함수	993.23	18
남녀별 3차함수	1025.43	21
남녀별 접목된 다항식	1062.22	23
최종 모형 ²⁾	1056.56	17

1) 0-4, 5-14, 30-44, 45-59, 60+

2) 남녀별 접목된 다항식을 사용한 모형에서 유의하지 않은 변수를 제외한 후에 구한 모형

반영하기 위한 성, 연령 변수의 선택은 중요하다고 생각된다. 남녀별 3차함수의 경우 Chi-Square 값이 상당히 높았는데 30대 연령까지는 접목된 다항식과 비슷한 형태를 보여주고 있기 때문인 것으로 여겨진다. 그러나 40대 연령 부터는 3차함수에 의해 잘 설명되지 않으므로 본 접목된 다항식 보다는 설명력이 떨어졌다.

4. 결 론

의료수요의 예측은 보건정책수립에 중요한 과제이다. 전국민 의료보험 실시에 따른 의료이용의 양적 증가와 아울러 인구 노령화에 따른 의료이용의 구조적 변화는 추후 의료수요 양상에 변화를 미칠 것으로 생각된다. 또한 노인의 의료이용율이 선진국과 같이 점차 증가할 것으로 생각되므로 인구학적 변수에 따른 의료이용율의 변화 양태에 대한 연구는 추후 의료정책 수립을 위해서도 중요한 것이다. 성, 연령별 의료이용율의 구조적 변화 및 양상은 앞으로 면밀히 관측, 검토되어야 하며 이를 위해서는 종래의 단순한 가변수 처리나 2, 3차 함수의 이용보다는 접목된 다항식의 이용이 바람직한 것으로 생각된다.

참고문헌

- 노공균. 의료수요의 가격탄력도에 관한 연구, 한국인구보건연구원, 1985
이현협. 보건의료 이용의 결정요인에 대한 경제학적 연구. 한국과학기술원 석사학위논문, 1982
한달선, 권순호, 권순원, 황성주. 춘천시민의 의료이용양상과 연관요인. 한림대 사회의학연구소, 1986
Lawless J.F., Statistical Models and Methods for LIFETIME Data. NY, John Wiley & Sons, 1982
Lee, K.S., Effects of Medical Insurance on the Demand for Medical Care in Korea, U. of Hawaii, 1984
McCullagh P., Nelder J.A., Generalized Linear Models, 2nd Edition, NY, Chapman and hall, 1989
Ohmura J., Analysis of Factors Affecting the Need and Demand for Medical Care, Social Science & Medicine, 1982 ; 12A : 485-496
Prenter P.M., Splines and Variational Methods. NY, John Wileys & Sons, 1975
Sard A., Weintraub S., A Book of Splines, John Wiley & Sons, 1971
Van der Gaag J. Van de Ven Wynand P.M.M., The Demand for Primary Health Care, Medical Care, 1978 ; 16 : 299-312