

## Build the nomogram by risk factors of chronic obstructive pulmonary disease (COPD)

Ju-Hyun Seo<sup>a,1</sup> · Dong-Yep Oh<sup>b,1</sup> · Yong-Soo Park<sup>c</sup> · Jea-Young Lee<sup>a,2</sup>

<sup>a</sup>Department of Statistics, Yeungnam University;

<sup>b</sup>Gyeongsangbuk-Do Livestock Research Institute;

<sup>c</sup>Department of Equine Industry, Korea National College of Agriculture and Fisheries

(Received June 14, 2017; Revised July 20, 2017; Accepted August 10, 2017)

---

### Abstract

The concentration of fine dust has increased in Korea and people have become more concerned with respiratory diseases. This study selected risk factors for chronic obstructive pulmonary disease (COPD) through demographic and clinical features and constructed a nomogram. First, logistic regression analysis was performed using demographic and clinical feature and the pulmonary function test results of the Korean National Health and Nutrition Examination Survey (KNHANES) 6<sup>th</sup> (2013–2015) and the nomogram was constructed to visualize the risk factors of chronic obstructive pulmonary disease in order to facilitate the interpretation of the analysis results. The ROC curve and calibration plot were also used to verify the nomogram of chronic obstructive pulmonary disease.

Keywords: chronic obstructive pulmonary disease (COPD), logistic regression analysis, nomogram, risk factors

---

### 1. 서론

최근 미세먼지의 농도가 점점 증가하는 추세이며 연구 발표에 따르면 미세먼지 노출은 폐기능 감소와 관련되어 있고, 성인에게는 폐기능 감소 속도 증가를 유발한다는 결과가 있어 사람들은 호흡기 계통 질환에 대하여 큰 관심을 가지고 있다 (Kyung 등, 2015). 본 논문에서 다루고 있는 만성 폐쇄성 폐질환(chronic obstructive pulmonary disease; COPD)은 유해한 입자나 가스의 흡입에 의해 폐에 염증 반응이 일어나면서 이로 인해 점막 기류 제한이 진행되어 폐 기능이 저하되고 호흡곤란을 유발하게 되는 호흡기 질환이다. 대부분의 사람들은 수년 동안 만성 기침이나 가래 생성과 같은 초기 증상을 무시하게 되고 이후 호흡 곤란이 발생할 정도로 폐기능이 손실되면 심할 경우 사망에 이르기까지 한다 (Zinlinsky와 Bednarek, 2001). 실제로 만성 폐쇄성 폐질환은 전세계적으로 사망 원인이 되는 질병 중

---

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through Agri-Bio industry Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (116065).

<sup>1</sup>These authors contributed equally to this work.

<sup>2</sup>Corresponding author: Department of Statistics, Yeungnam University, 280 Daehak-Ro, Gyeongsan, Gyeongbuk 38541, Korea. E-mail: [jlee@yu.ac.kr](mailto:jlee@yu.ac.kr)

높은 순위를 차지하고 있고, 발병의 위험 요인으로는 유전적 요인, 흡연, 내·외부적 공기 오염, 나이, 천식, 성별 등이 있다고 발표된 바 있다 (Mannino, 2007). 우리나라의 경우에는 40세 이상 폐쇄성 폐질환 유병률이 2009년, 2015년에 각각 10.5%, 12.3%인 것으로 보고되었다 (Korea Centers for Disease Control and Prevention, 2016). 또한 만성 폐쇄성 폐질환을 포함한 호흡계통의 질환으로 인한 사망률이 인구 10만 명당 54.6명인 것으로 나타났다 (Korean Statistical Information Service, 2015). 이처럼 국·내외에서 사망의 원인으로 중요하게 다뤄지는 질병인 만큼 만성 폐쇄성 폐질환의 위험 요인을 선별하여 이를 인지한 후 발병률을 줄일 수 있는 치료법을 개발하는 것이 가장 중요하다.

일반적으로 질병의 위험 요인을 선별하기 위한 분석 방법에는 로지스틱 회귀분석과 Cox의 비례위험모형이 많이 쓰인다. 이러한 분석 방법을 통해 최근 국민건강영양조사 4기와 5기 자료를 토대로 만성 폐쇄성 폐질환의 위험 요인을 선별하기 위한 연구가 진행되었고, 선별된 위험 요인으로는 연령, 성별, 거주 지역, 교육 정도, 흡연량, 천식, 결핵 등이 있다는 결과를 보였다 (Jung 과 Lee, 2011; Park 등, 2015). 하지만, 통계학을 전문적으로 배우지 않은 의료종사자와 같은 비전문가의 경우 이러한 분석 방법을 해석하는 것에는 한계가 있다. 이러한 부분을 보완하기 위해 만들어진 통계학적 도구가 바로 노모그램이다. 노모그램은 의사결정을 지원하는데 사용되며 예측한 확률에 대하여 직접적인 시각적 설명을 제공해준다 (Mozina 등, 2004). 해외에서는 이미 다양한 질병들에 대한 노모그램이 개발되었고, 국내에서도 위암, 전립선암, 침윤암, 골육종 등의 질병에 대한 노모그램이 개발되었다 (Jun, 2015; Lee와 Chang, 2014; Ahn, 2013; Kim 등, 2014). 현재 개발되어 있는 노모그램은 대체로 암과 같은 심각한 질병 위주로 많이 개발되어 있지만, 한 번 발병하면 치료가 어려운 만성적 질병의 노모그램 개발은 아직 이루어지지 않아서 그 필요성을 인식하였다. 따라서 본 연구에서는 로지스틱 회귀분석을 통해 만성 폐쇄성 폐질환의 위험 요인을 선별하고 이를 시각화하는 노모그램을 구축한다.

본 연구는 다음과 같은 순서로 구성되어 있다. 2절에서는 로지스틱 회귀분석과 노모그램에 대해 소개하고 3절에선 만성 폐쇄성 폐질환의 위험요인을 선별하기 위해 사용된 자료와 인구학적 및 임상적 특징을 설명한다. 4절에서는 2절에서 소개된 분석 방법을 이용하여 COPD의 분석 결과를 해석하고 만성 폐쇄성 폐질환에 대한 노모그램을 구축한다. 마지막으로 5절에서는 분석의 결과와 발전 방향을 제시한다.

## 2. 분석방법

### 2.1. 로지스틱 회귀분석(logistic regression)

종속변수가 연속형 변수일 경우에는 일반적인 회귀분석을 사용하지만, 종속변수가 성공/실패, 생존/사망, 질병의 유무 등의 범주로 표현되는 이산형 변수일 경우에 로지스틱 회귀분석을 사용한다. 로지스틱 회귀분석은 병리학 연구에서 질병과 연관 있는 위험 요인들을 식별하거나, 임상 연구 자료에서 중요한 요인들을 식별하는 탐색적 분석에 많이 적용 된다 (Lee 등, 2005; Heo와 Lee, 2008). 독립변수는 연속형, 이산형인 경우 둘 다 가능하며 정규분포를 가정하지 않는다.

본 연구에서는 질병의 유무가 종속변수이므로 이항반응에 대해서 살펴보기로 한다. 이항반응변수는 베르누이 변수라고도 하고 이 변수에 대한 분포는 성공에 대한 확률  $P(Y = 1) = p_x$ 와 실패에 대한 확률  $P(Y = 0) = 1 - p_x$ 로 표현된다. 즉, 반응변수  $Y$ 는 설명변수가  $X$ 일 때의 성공확률  $p_x$ 를 갖는 분포를 따른다고 가정할 때, 설명변수  $X = x$ 일 때 반응변수  $Y = y$ 일 확률은

$$P(Y = y|X = x) = p_x^y(1 - p_x)^{1-y}, \quad (y = 0, 1)$$

이 된다. 이때 성공확률  $p_x$ 에 대해 다음과 같은 선형 확률모형을 생각할 수 있다.

$$p_x = \alpha + \beta x.$$

하지만 위의 선형 확률 모형은 구조적 단점을 가진다. 먼저  $p_x$ 의 값이 (0, 1)의 범위를 가지고  $\alpha + \beta x$ 는 실수 전체의 값을 가지므로 성공확률  $p_x$ 가 범위 외의 값을 가질 수 있게 된다. 또한 최소제곱추정량은 분산  $\text{Var}(Y) = p_x(1 - p_x)$ 에서 분산이  $x$ 의 영향을 받기 때문에 분산이 일정하지 않아 더 이상 최소 분산을 가지지 못한다. 따라서 위의 식이 아닌 성공확률  $p_x$ 에 대해  $k$ 개의 설명변수  $X_1, X_2, \dots, X_k$ 와 비선형인 아래의 식과 같은 함수를 생각하게 된다.

$$p_x = \frac{\exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}{1 + \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}$$

성공확률  $p_x$ 를 세로축으로 하고 설명변수  $X_1, X_2, \dots, X_k$ 를 가로축으로 했을 때의 그래프 모양은 s자 곡선의 형태를 갖는다.  $|\beta|$ 가 커질수록 곡선의 변화율도 커지게 된다. 위의 식에서 성공확률에 대한 오즈(odds)는 다음과 같고,

$$\frac{p_x}{1 - p_x} = \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

성공확률의 오즈에 로그를 취한 로그 오즈는 아래의 선형식으로 나타낼 수 있다.

$$\ln\left(\frac{p_x}{1 - p_x}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

위 식을 로지스틱 회귀모형이라 한다.

따라서 본 연구는 만성 폐쇄성 폐질환의 여부를 반응변수로 하고 단계별 선택 방법을 사용하여 로지스틱 회귀분석을 실시하였고 만성 폐쇄성 폐질환의 위험 요인들을 선별하였다.

## 2.2. 노모그램(nomogram) 구축

노모그램은 특정한 결과에 대한 가능성을 예측하고자 고안된 통계학적인 도구로써 특정 질병과 관련된 위험 요인들뿐만 아니라 환자의 특성을 바탕으로 구축된다 (Jun, 2015). 따라서 보다 단순한 그래픽 표현을 사용하여 읽기 쉽고, 복잡한 계산 없이 결과 확률을 예측하는 데 사용할 수 있다 (Mozina 등, 2004).

노모그램은 추정된 회귀계수들로부터 계산된 각 위험 요인들의 영향력을 점수화하여 나타내주는 points 선, 위험 요인에 할당된 각각의 점수의 누적 합을 나타내주는 total points (TP) 선, TP로부터 대응되는 결과 변수의 확률을 나타내는 probability 선으로 구성되어 있다. 각각의 위험 요인들은 추정된 회귀계수 값을 이용하여 0부터 100점까지의 점수로 표현된다. 추정된 회귀계수들의 절댓값 중 가장 큰 값을 가지게 되는 요인이 Points 100점을 받게 되고, 이는 질병 발생에 가장 큰 영향을 미친다는 것을 의미한다. 일반적으로 병의 발생 여부나 재발 여부와 같이 결과변수가 이분형인 경우 로지스틱 회귀모형을 구축하고, time-to-event인 경우 Cox 비례위험모형을 구축한다. 본 연구에서 결과변수는 만성 폐쇄성 폐질환의 여부를 나타내는 이분형이므로 로지스틱 회귀분석을 사용하여 노모그램을 개발하였다. 노모그램의 구체적인 개발 절차는 다음과 같다 (Iasonos 등, 2008; Yang, 2014).

### (1) 위험 요인 선택

교차분석(카이제곱검정)을 사용하여 위험 요인을 선택한다.

### (2) 선택된 모형(로지스틱 회귀모형)으로 분석

Hosmer-Lemeshow 검정을 이용하여 로지스틱 회귀모형의 적합도를 확인한다.

### (3) 노모그램 구축

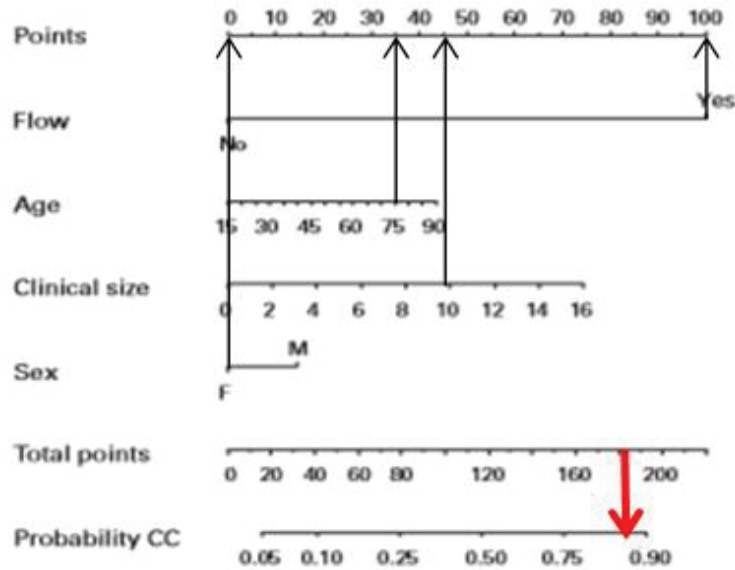


Figure 2.1. Nomogram plot (Iasonos *et al.*, 2008).

- 각 위험요인에 할당된 점수를 나타내는 Points 선 표현

각 위험 요인의 Points 값은 0에서 100 사이의 값을 가지게 된다. 로지스틱 회귀모형으로부터 구한 추정 회귀계수 값을 이용하여 각 위험 요인별로 linear predictor 값( $LP_{ij}$ )을 계산한다. 이때,  $i = 1, 2, \dots, m$ 는 위험요인의 수,  $j = 1, 2, \dots, n_i$ 는 각 위험 요인 당 범주 수이다.

$$LP_{ij} = \beta_i \times X_{ij},$$

$$\text{Point}_{ij} = \frac{LP_{ij} - \min LP_{ij}}{\max LP_{*j} - \min LP_{*j}} \times 100.$$

$LP_{ij}$  값을 이용하여 각 위험 요인별로 점수(points)를 계산한다.  $LP_{*j}$ 는 추정 회귀계수의 절댓값이 가장 큰 요인의  $LP_{ij}$  값이라 할 수 있다. 계산된 값을 통해 Figure 2.1에서 각 위험 요인들의 영향력을 나타내주는 선이 된다. 이 선의 길이가 길수록 특정 질병에 영향을 미치는 정도가 크다고 할 수 있다.

- 위험 요인에 할당된 각각의 점수의 누적 합을 나타내는 TP 선

$$\text{LP값 당 단위 점수(points per unit of linear predictor)} : \frac{100}{\max LP_{*j} - \min LP_{*j}}.$$

- ①  $LP_{\text{for TP}=0}$  : TP가 0의 경우 LP값 (상수값)
- ②  $LP_{\text{for TP}>0}$  :  $\ln \left( \frac{P(\text{risk of } Y = 1)}{1 - P(\text{risk of } Y = 1)} \right)$

$$\begin{aligned} \text{TP} &= \text{LP값 당 단위 점수} \times (\text{②} - \text{①}) \\ &= \left( \frac{100}{\max LP_{*j} - \min LP_{*j}} \right) \times (\text{②} - \text{①}). \end{aligned}$$

위의 식과 같이 LP값 당에 해당되는 단위 점수를 구하고 계산된 LP값의 곱으로 TP를 계산할 수 있다. Figure 2.1에서 TP 선으로 나타나는 값이다.

- TP에 대응되는 결과 변수의 확률을 나타내는 Probability 선

임의의 TP에 대응되는 확률 값을 나타내는 선이며, 이 값은 임의의 TP로 변환된  $LP_{for TP>0}$  값을 이용하여 로지스틱 회귀모형 식을  $P(\text{risk of } Y = 1)$ 에 대한 식으로 정리하면 구할 수 있다.

위 식들을 이용하여 계산한 점수들로 Figure 2.1과 같이 노모그램을 구축할 수 있다. 개발된 노모그램에 따라 특정 질병에 대해 환자 개개인의 발병률이나 생존율을 예측하는데 사용될 수 있다. Figure 2.1의 노모그램을 살펴보면 4개의 변수 중 Flow의 point 선이 100점으로 가장 길기 때문에 영향을 미치는 정도가 가장 크다고 할 수 있다. 또한 Age는 45점, Clinical size는 75점, Sex는 15점을 최대로 얻을 수 있다. 예를 들어 한 환자에 대한 결과가 Flow(Yes), Age(75), Clinical size(10), Sex(F)일 경우 TP는 180점이 되고 이에 따라 특정 질병의 발병률이 약 90%가 됨을 확인할 수 있다.

(4) 노모그램에 대한 검증

로지스틱 회귀모형을 사용하여 구축한 노모그램에 대한 검증은 Receiver operation characteristic (ROC) 곡선 아래의 면적(area under the ROC curve; AUC)과 Calibration plot을 통하여 이루어진다.

① Discrimination: AUC

ROC 곡선은 진단 및 예후 모델의 임상적 유용성을 평가하기 위해 사용되어 진다 (Cook, 2008). ‘민감도’를 수직축, ‘1-특이도’를 수평축으로 하여 연결선을 그려내고, ROC 곡선이 대각선 위쪽에 위치하게 될수록 좋은 성능을 가진 모형이라고 할 수 있다. AUC가 예측모형의 성능을 측정하는 값으로 사용될 수 있는데, 이 값은 0.5와 1사이에 존재하며 면적이 넓은 만큼 값이 1에 가까워지므로 이는 예측 모형의 성능이 좋음을 나타낸다. 따라서 노모그램의 유용성을 평가하기 위해 AUC를 이용하였다.

② Calibration: Calibration plot

Calibration plot은 노모그램에서 예측한 확률과 관찰된 확률이 얼마나 일치하는가를 확인하는 것이다 (D’Agostino 등, 2001; Nam과 D’Agostino, 2002). 노모그램을 통해 예측한 확률과 실제 관찰된 확률이 완벽하게 정확할 때 45° 각도로 그려진 선이 나타나게 되므로, 예측한 확률과 관찰된 확률에 따른 선이 45° 각도 선에 가까울수록 정확한 예측력을 보인다고 할 수 있다 (Iasonos 등, 2008). 따라서 노모그램에서 예측한 발병률과 실제 발병률이 얼마나 일치하는지를 검증하기 위해 Calibration plot을 이용하였다.

### 3. 연구자료와 특징

분석을 위하여 본 연구는 국민건강영양조사(Korean National Health and Nutrition Examination Survey; KNHANES) 6기(2013–2015) 자료를 사용하였다. 국민건강영양조사는 국민의 건강수준, 건강행태, 식품 및 영양섭취 실태에 3가지 조사를 걸쳐 국가 및 시도 단위의 대표성과 신뢰성을 갖춘 통계를 산출하기 위해 시행되었다. 조사 참여자 22,948명 중에서 폐기능 검사를 실시한 40세 이상의 응답자를 대상으로 하였고, 결측치가 포함되어 있는 응답자는 제외하여 총 8,258명으로 분석을 진행하였다.

만성 폐쇄성 폐질환의 여부는 폐기능 검사 결과에 따라 범주화 되었다. Global Initiative for Chronic Obstructive Lung Disease (GOLD)의 지침에 따르면 폐기능 검사를 통하여  $FEV_1/FVC < 0.7$ 일 경

**Table 4.1.** Chi-square test results for 12 risk factors

Risk factors		COPD(%)	Non-COPD(%)	<i>p</i> -value
Sex	Men	1964 (77.2)	580 (22.8)	0.000
	Women	3040 (93.9)	197 (6.1)	
Age	< 65	3848 (92.1)	328 (7.9)	0.000
	≥ 65	1156 (72.0)	449 (28.0)	
Residence	Urban	4047 (87.4)	586 (12.6)	0.000
	Rural	957 (83.4)	191 (16.6)	
Education	≤ Middle	1954 (82.0)	430 (18.0)	0.000
	High school	1713 (88.7)	218 (11.3)	
	≥ University	1337 (91.2)	129 (8.8)	
Smoking	No	3249 (92.8)	251 (7.2)	0.000
	Yes	1755 (76.9)	526 (23.1)	
Hypertension	No	3674 (89.1)	449 (10.9)	0.000
	Yes	1330 (80.2)	328 (19.8)	
Ischemic heart disease	No	4893 (86.8)	743 (13.2)	0.000
	Yes	111 (76.6)	34 (23.4)	
Tuberculosis	No	4801 (87.3)	707 (12.7)	0.000
	Yes	203 (72.8)	76 (27.2)	
Asthma	No	4904 (87.4)	704 (12.6)	0.000
	Yes	100 (57.8)	73 (42.2)	
Diabetes	No	4514 (87.4)	651 (12.6)	0.000
	Yes	490 (79.5)	126 (20.5)	
Depression	No	4729 (86.3)	748 (13.7)	0.040
	Yes	275 (90.5)	29 (9.5)	
Stroke	No	4904 (86.8)	749 (13.2)	0.005
	Yes	100 (78.1)	28 (21.9)	

우 만성 폐쇄성 폐질환 환자라고 판단하게 된다. 이때, FEV<sub>1</sub>은 1초간 노력성 호기량이고, FVC는 노력성 폐활량을 말한다. 위험 요인으로는 인구학적 특징인 성별, 나이, 거주지, 교육수준, BMI지수, 소득분위와 임상적 특징인 흡연 여부, 동반질환(고혈압, 협심증, 류마티스 관절염, 폐결핵, 천식, 당뇨, 폐암, 우울증, 뇌졸중)을 사용하였다. 나이는 65세 이상과 65세 미만으로 2개의 범주로, 거주지는 동과 읍·면으로 2개의 범주로 나누었다. 소득분위는 소득 사분위수 구분 기준금액에 따라 4개의 범주로 나누었다. BMI지수는 25를 기준으로 비만 여부를 판단하고 있다. 따라서 < 25와 ≥ 25로 범주를 나누었다. 흡연 여부는 일생동안 5갑(100개비) 미만을 피웠거나 피운 적 없는 사람을 비흡연자, 5갑(100개피) 이상 피운 사람을 흡연자로 정의하여 범주를 나누었다. 마지막으로 동반질환의 경우 이전에 특정 질병을 의사에게 진단 받은 적이 있는지에 대한 답변을 이용하여 동반질환의 유무를 판단하였다.

#### 4. COPD의 분석 결과

본 연구는 국민건강영양조사(KNHANES) 6기(2013–2015) 자료를 사용하여 폐기능 검사를 실시한 40세 이상의 대상자로 분석을 실시하였다. 결측값을 가지고 있는 대상자는 분석에서 제외되었다. 그리고 로지스틱 회귀분석의 모형을 만들고 노모그램을 구축하여 검증하기 위해서 자료를 무작위로 Training data( $n = 5781$ )와 Validation data( $n = 2477$ )로 나누어서 분석을 Training data로 노모그램을 구축하고 Validation data로 검증하였다.

**Table 4.2.** Multiple logistic regression analysis result using the 12 risk factors

Risk factors		Odd ratio	95% CI	p-value
Sex	Men	3.771	2.883–4.933	0.000
	Women	1	.	
Age	< 65	1	.	0.000
	≥ 65	3.942	3.297–4.713	
Education	≤ Middle	2.053	1.614–2.611	0.000
	High school	1.480	1.156–1.895	0.002
	≥ University	1	.	
Smoking	No	1	.	0.000
	Yes	1.872	1.459–2.400	
Tuberculosis	No	1	.	0.000
	Yes	1.953	1.428–2.671	
Asthma	No	1	.	0.000
	Yes	7.205	5.013–10.356	

p-value of Hosmer and Lemeshow goodness-of-fit test is 0.376.

#### 4.1. 교차분석

위험 요인으로 선별된 인구학적, 임상적 특징 16가지를 만성 폐쇄성 폐질환의 여부와 독립적으로 관련이 있는지 확인하기 위해 교차분석을 통하여 알아보았다. 결과를 통하여 총 12가지 위험 요인이 유의한 결과를 보임을 확인하였다 (Table 4.1).

#### 4.2. 로지스틱 회귀분석

만성 폐쇄성 폐질환의 위험 요인들을 식별하기 위해 로지스틱 회귀분석을 실시하였다. Table 4.2는 교차분석을 통해 유의한 결과를 보였던 12개의 특징을 단계별 선택 방법(stepwise regression)을 이용하여 로지스틱 회귀분석을 실시한 결과이다. Hosmer-Lemeshow 적합도 검정(Hosmer-Lemeshow goodness-of-fit test)은 유의확률 0.376으로 적합하다는 결과를 얻었다. 결과적으로 최종 선별된 위험 요인은 성별(sex), 나이(age), 교육수준(education), 흡연(smoking), 폐결핵(tuberculosis), 천식(asthma)이다.

#### 4.3. COPD 노모그램 구축

노모그램은 특정 질병에 영향을 주는 위험 요인을 한눈에 알아볼 수 있을 뿐만 아니라 영향력까지 표현이 가능하다. 또한 영향력을 점수화하여 환자의 특정 질병 발병률을 예측할 수 있다. 따라서 본 연구에서는 만성 폐쇄성 폐질환의 위험 요인을 단계별 선택 방법을 이용한 로지스틱 회귀분석을 통하여 선별하였고 발병률을 예측하는 노모그램을 구축하였다 (Figure 4.1).

COPD의 노모그램을 보면 Points를 나타내는 선 아래로 위험 요인으로 선별된 성별, 나이, 교육 수준, 흡연 여부, 폐결핵, 천식 각각의 회귀계수를 점수화하여 선이 표현되어져 있다. 회귀계수 값이 커질수록 점수화된 값 또한 커지게 되고 이에 따라 각 선의 길이가 결정된다. 그리고 TP 선은 위험 요인에 할당된 각각의 점수를 합한 값을 나타내주며 이 값에 따라 Probability 값을 구해낼 수 있게 된다. Figure 4.1을 보면 위험 요인 중 천식이 가장 길이가 긴 것을 확인할 수 있고 이는 만성 폐쇄성 폐질환의 발병에 가장 큰 영향을 주는 요인임을 알 수 있다. 그 다음으로 나이, 성별, 교육수준, 결핵, 흡연 여부 순으로 발병에 영향을 미쳤다. 예를 들어 70세 남자이면서 학력이 고졸이고 흡연자이면서 천식을 앓는 사람이

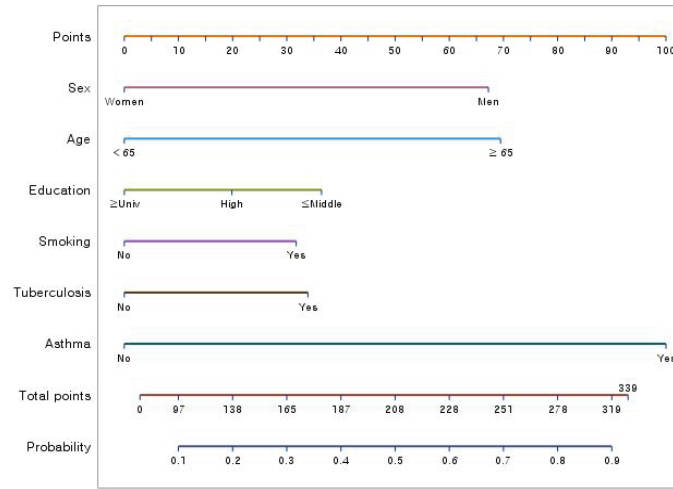


Figure 4.1. Nomogram about chronic obstructive pulmonary disease.

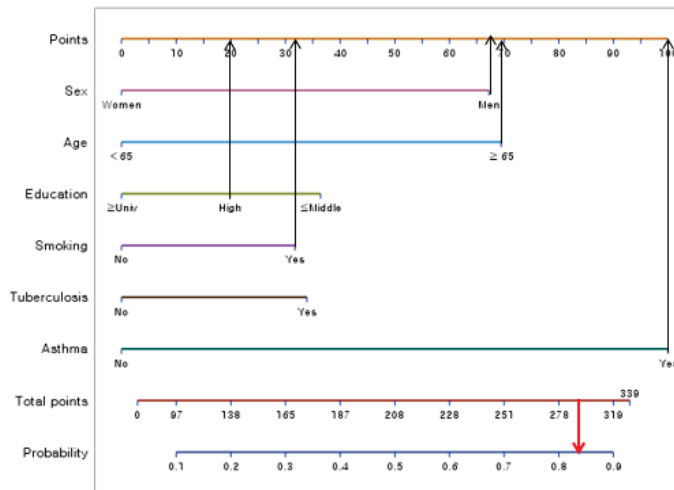


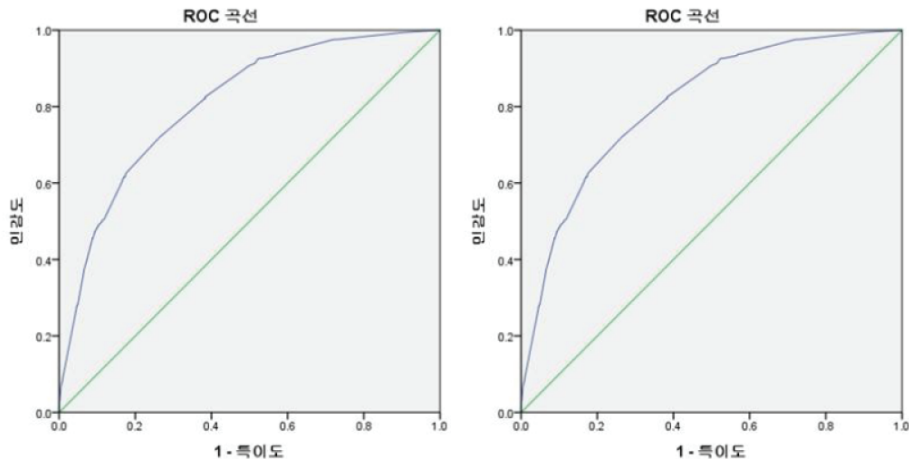
Figure 4.2. Incidence rate calculation example using nomogram.

라면 TP가 288.28으로 만성 폐쇄성 폐질환의 발병률은 83%라 예측할 수 있다.

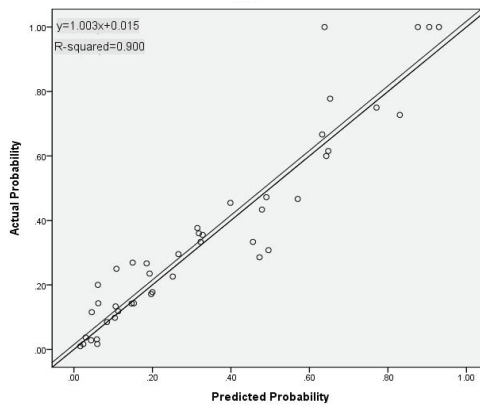
#### 4.4. 노모그램 검증

- ① Discrimination: AUC, ROC 곡선은 진단 및 예후 모델의 임상적 유용성을 평가하기 위해 사용되어 진다 (Cook, 2008). ‘민감도’를 수직축, ‘1-특이도’를 수평축으로 하여 연결선을 그려 작성할 수 있으며, ROC 곡선이 대각선 위쪽에 위치하게 될수록 좋은 성능을 가진 모형이라고 할 수 있다. Figure 4.3의 그래프를 보면 Training data를 사용하여 그려진 ROC 곡선의 AUC는 0.807( $p = 0.000$ )이며, Validation data로 그려진 ROC 곡선의 AUC는 0.811( $p = 0.000$ )로 통계적으로 유의하다는 것을 알 수 있었다.





(a) Training data (b) Validation data  
**Figure 4.3.** Receiver Operation Characteristic (ROC) curve.



**Figure 4.4.** Calibration plot.

② Calibration: Calibration plot은 노모그램에서 예측한 확률과 관찰된 확률이 얼마나 일치하는가를 확인하는 것으로 (D’Agostino 등, 2001; Nam과 D’Agostino, 2002), Figure 4.4는 노모그램의 예측 확률에 따라서 44개의 그룹으로 환자들을 나누고, 각 그룹의 평균값을 이용하여 관찰 확률과 비교해 본 Calibration plot이다. Calibration plot의 결정계수( $R^2$ )는 0.900로 추정된 직선이 그려졌으므로 구축된 노모그램이 만성 폐쇄성 폐질환의 발병률을 예측하는데 적합하다고 판단된다.

**5. 결론 및 토의**

본 연구에서는 만성 폐쇄성 폐질환의 위험 요인을 선별하기 위하여 로지스틱 회귀분석을 실시하였고 이를 시각화하는 노모그램을 구축하였다. 노모그램을 검증하기 위해 자료를 무작위로 Training data( $n = 5781$ )와 Validation data( $n = 2477$ )로 나누어서 Training data의 분석을 통해 노모그램을 구축하고 Validation data로 검증하였다. 사용된 데이터는 국민건강영양조사(KNHANES) 6기(2013-2015) 자

료를 사용하였고, 연구 대상자는 40세 이상의 폐기능 검사를 실시한 사람을 기준으로 선별되었다. 총 5,781명 중 777명이 만성 폐쇄성 폐질환이라 판단되었고, 인구학적 특징인 성별, 나이, 거주지, 교육 수준, BMI 지수, 소득 분위와 임상적 특징인 흡연 여부, 동반 질환(고혈압, 협심증, 류마티스 관절염, 폐결핵, 천식, 당뇨, 폐암, 우울증, 뇌졸중)을 사용하여 위험 요인을 선별하였다. 16가지 특징을 만성 폐쇄성 폐질환의 여부와 독립적으로 관련이 있는지 확인하기 위해 교차분석을 실시하였다. 그 결과 총 12가지 위험 요인(성별, 나이, 거주지, 교육 수준, 흡연 여부, 고혈압, 협심증, 폐결핵, 천식, 당뇨, 우울증, 뇌졸중)이 유의한 결과로 나타났으며, 유의한 결과를 보였던 12개의 위험 요인을 사용하여 단계별 선택 방법으로 로지스틱 회귀분석을 실시한 결과 성별, 나이, 교육 수준, 흡연 여부, 폐결핵, 천식이 유의한 결과를 보였다. Maninno (2007)에서의 COPD 위험 요인들과 매우 흡사한 모습을 보였으며, 이러한 위험 요인들을 이용하여 노모그램을 구축하였다. 만성 폐쇄성 폐질환의 노모그램을 보면 천식이 100점을 할당받아 만성 폐쇄성 폐질환의 발병에 가장 큰 영향을 주는 요인이 되며, 성별과 나이의 선 길이가 비슷하므로 천식 다음으로 큰 영향을 주는 요인이 된다. 또한 교육 수준, 흡연여부, 폐결핵의 선 길이가 비슷하여 앞서 3개의 요인보다 상대적으로 영향을 덜 끼친다고 볼 수 있다. 구축된 노모그램을 검증하기 위해 ROC curve를 그려보았을 때 Training data와 Validation data의 AUC가 각각 0.807( $p = 0.000$ ), 0.811( $p = 0.000$ )로 통계적으로 유의하다는 것을 알 수 있었다. 그리고 노모그램의 예측 확률에 따라서 44개의 그룹으로 환자들을 나누고, 각 그룹의 평균값을 이용하여 관찰 확률과 비교하는 Calibration plot을 그린 결과, 결정계수( $R^2$ )가 0.900로 추정된 직선이 그려졌으므로 구축된 노모그램이 만성 폐쇄성 폐질환의 발병률을 예측하는데 적합하다고 판단되었다.

위 결과는 Figure 4.1의 노모그램을 통해서 나타났으며 이에 따라 비전공자인 의료종사자들은 분석 결과를 좀 더 쉽게 해석할 수 있다. 또한 일반인들도 구축된 노모그램 점수화된 값을 계산할 수 있으며, 만성 폐쇄성 폐질환의 발병률을 복잡한 계산 없이 쉽게 예측할 수 있다. 이러한 분석 결과를 통해 만성 폐쇄성 폐질환의 위험 요인에 따라 발병률을 미리 예측하여 알맞은 예방 처치를 하고 질병에 맞는 치료법을 발전시키는 것이 필요하다고 본다.

## References

- Ahn, J. H. (2013). *Nomogram for prediction of prostate cancer in Korean men with serum prostate-specific antigen less than 10ng/mL*, Busan University, Busan.
- Cook, N. R. (2008). Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve, *Clinical Chemistry*, **54**, 17–23.
- D'Agostino, R. B., Grundy, S., Sullivan, L. M., and Wilson, P. (2001). Validation of the Framingham coronary heart disease prediction scores, *Journal of the American Medical Association*, **286**, 180–187.
- Heo, M. H. and Lee, Y. G. (2008). *Data Mining Modeling and Example*, Hannarae, Seoul.
- Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to build and interpret a nomogram for cancer prognosis, *Journal of Clinical Oncology*, **26**, 1364–1370.
- Jun, H. J. (2015). Establishment of a nomogram to predict the prognosis of metastatic or recurrent gastric cancer patients (master's thesis), Yonsei University, Seoul.
- Jung, Y. M and Lee, H. Y. (2011). Chronic obstructive pulmonary disease in Korea: prevalence, risk factors, and quality of life, *Journal of Korean Academy of Nursing*, **41**, 149–156.
- Kim, S. H., Shin, K. H., Kim, H. Y., Cho, Y. J., Noh, J. K., Suh, J. S., and Yang, W. I. (2014). Postoperative nomogram to predict the probability of metastasis in Enneking stage IIB extremity osteosarcoma, *BMC Cancer*, **14**, 666.
- Korea Centers for Disease Control and Prevention (2013–2015). The Sixth National Health and Nutrition Examination Survey (KNHANES VI), from: <http://knhanes.cdc.go.kr/>
- Korea Centers for Disease Control and Prevention (2016). Korea Health Statistics 2015 : Korea National Health and Nutrition Examination Survey (KNHANES VI-3), Cheongju, from: <https://knhanes.cdc.go>

- kr/knhanes/sub04/sub04.03.do? bclassType=7
- Korean Statistical Information Service (2015). Cause of Death, from: [http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT\\_1B34E01&conn\\_path=I2](http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1B34E01&conn_path=I2)
- Kyung, S. Y., Kim, Y. S., Kim, W. J., Park, M. S., Song, J. W., Yum, H. K., Yoon, H. K., Rhee, C. K., and Jeong, S. H. (2015). Guideline for the prevention and management of particulate matter/Asian dust particle induced adverse health effect on patients with pulmonary diseases, *Journal of the Korean Medical Association*, **58**, 1060–1069.
- Lee, J. W., Park, M. R., and Yu, H. N. (2005). *Statistical Method for Bioscience Research*, Freedom Academy, Seoul.
- Lee, S. C. and Chang, M. C. (2014). Development and validation of web-based nomogram to predict post-operative invasive component in ductal carcinoma in situ at core needle breast biopsy, *Healthcare Informatics Research*, **20**, 152–156.
- Mannino, D. M. (2007). Global burden of COPD: risk factors, prevalence, and future trends, *The Lancet*, **370**, 765–773.
- Mozina, M., Demsar, J., Smrke, D., and Zupan, B. (2004). Nomograms for Naive Bayesian classifiers and how can they help in medical data analysis, *MEDINFO 2004*, **1762**, 765–773.
- Nam, B. H. and D'Agostino, R. B. (2002). Discrimination index, the area under the ROC curve, In *Goodness-of-Fit Tests and Model Validity*, 267–279.
- Park, H. Y., Jung, S. Y., Lee, K. H., Bae, W. K., Lee, K. H., Han, J. S., Kim, S. R., Choo, S. Y., Jeong, J. M., Kim, H. R., Ro, H. J. and Jeong, H. S. (2015). Prevalence of chronic obstructive lung disease in Korea using data from the fifth Korea National Health and Nutrition Examination Survey, *Korean Journal of Family Medicine*, **36**, 128–134.
- Yang, D. (2014). Build prognostic nomograms for risk assessment using SAS. In *Proceedings of SAS Global Forum 2013*, from: <http://support.sas.com/resources/papers/proceedings13/264-2013.pdf>.
- Zinlinsky, J. and Bednarek, M. (2001). Early detection of COPD in a high-risk population using spirometric screening, *Chest*, **119**, 731–736.

# 만성 폐쇄성 폐질환의 위험요인 선별을 통한 노모그램 구축

서주현<sup>a,1</sup> · 오동엽<sup>b,1</sup> · 박용수<sup>c</sup> · 이제영<sup>a,2</sup>

<sup>a</sup>영남대학교 통계학과, <sup>b</sup>경상북도축산기술연구소, <sup>c</sup>국립한국농수산대학 말산업학과

(2017년 6월 14일 접수, 2017년 7월 20일 수정, 2017년 8월 10일 채택)

## 요약

최근 미세먼지 농도가 올라감에 따라 사람들은 호흡기 질환에 큰 관심을 가지고 있다. 본 연구는 인구학적 및 임상적 특징을 통한 만성 폐쇄성 폐질환(chronic obstructive pulmonary disease)의 위험요인을 선별하고 이에 따른 노모그램을 구축하였다. 먼저 국민건강영양조사(KNHANES) 6기 (2013-2015)의 인구학적 및 임상적 특징, 폐기능 검사 결과를 사용하여 로지스틱 회귀분석을 실시 하였고 비전공자들도 분석 결과에 대한 해석을 쉽게 할 수 있도록 만성 폐쇄성 폐질환의 위험 요 인을 시각화한 노모그램을 구축하였다. 또한 ROC curve와 Calibration plot을 이용하여 만성 폐쇄 성 폐질환의 노모그램을 검증하였다.

주요용어: 만성 폐쇄성 폐질환, 로지스틱 회귀분석, 노모그램, 위험 요인.

본 연구는 농림축산식품부의 재원으로 농림수산식품기술기획평가원의 첨단생산기술개발사업의 지원을 받아 연구되었음(116065).

<sup>1</sup>공동저자.

<sup>2</sup>교신저자: (38541) 경상북도 경산시 대학로 280, 영남대학교 통계학과. E-mail: jlee@yu.ac.kr