

대한물리치료과학회지

Journal of Korean Physical Therapy Science
2024. 09. Vol. 31, No 3, pp. 31-50

한국인 폐활량 예측산식을 위한 예비타당성 연구: 통계검정모델 중심

이명모¹ · 오윤중¹ · 박삼호² · 강위창³

¹대전대학교 물리치료학과 · ²광주여자대학교 물리치료학과 · ³대전대학교 빅데이터학과

Preliminary Feasibility Study for Korean Lung Capacity Prediction Formula: Focused on Statistical Test Model

Myungmo Lee, Ph.D., P.T.¹ · Younjung Oh, MS, P.T.¹ · Samho Park, Ph.D, P.T.² · Weechang Kang, Ph.D.³

¹Dept. of Physical therapy, Daejeon University

²Dept. of Physical therapy, Kwangju Women's University

³Dept. of Statistics, Daejeon University

Abstract

Background: The lung capacity prediction formula in Korea is an important judgment standard. Since there is no appropriate lung capacity prediction formula, various prediction formulas are used for foreigners such as Northeast Asians. The purpose of this study is to develop a lung capacity prediction equation by selecting data and setting the selection criteria for normal subjects in accordance with international standards through strict quality control, and to propose a new prediction model.

Design: Preliminary feasibility study

Methods: A total of 857 people who met the criteria for normal people were finally collected. The tester used for the lung capacity test was the V-Max Encore 22 (Carefusion, California, USA), which is a lung capacity tester proposed by the Korean Society of Tuberculosis and Respiratory Medicine and satisfies accuracy and precision. Among the indicators measured using spirometry, forced vital capacity (FVC), forced expiratory volume in 1 second (FEV₁), forced expiratory volume ratio in 1 second (FEV₁/FVC), forced mid-expiratory flow (Forced expiratory flow 25-75%, FEF_{25-75%}) and peak expiratory flow (PEF) values were collected.

Results: This study confirmed a significant correlation between age, height, weight, and pulmonary function indicators. Additionally, it found a correlation between

body mass index, which considers the diversity of physical conditions, and pulmonary function indicators. Graphs depicting age-specific pulmonary function indicators by gender, presented as generalized additive model results from collected data, showed a pattern where both FVC and FEV₁ increased until the mid-20s and then gradually decreased with aging. FEV₁% and PEF exhibited a continuous decrease with aging.

Conclusion: This study confirms that there is a significant correlation between weight and pulmonary function in the prediction formula for lung capacity. Additionally, it verifies the correlation between body mass index, which considers the diversity of physical conditions, and pulmonary function. The study suggests that the predicted values are relatively low due to factors such as aging and environmental influences like COVID-19. This preliminary study holds clinical significance for improving the diagnostic accuracy of respiratory symptoms in the elderly.

Key words: Spirometry, Predictive Value of Tests, Lung Volume Measurements

교신저자

이명모

대전광역시 동구 대학로 62 응용과학과 2506호

T: 042-280-4295, E: mmlee@dju.ac.kr

I. 서론

폐활량 검사는 흡기와 호기 시 공기의 양을 유량과 유속의 특성을 고려하여 측정하는 검사로 폐의 기능적인 부분을 객관화한 자료로 평가하는 가장 대표적인 검사방법이다(Ranu 등 2011). 또한 검사 시간이 짧고, 검사 방법이 비교적 간단하여 중소병원에서도 시행할 수 있는 장점이 있다. 폐활량 검사는 일반 호흡기 질환의 진단 및 경과 관찰, 수술 후 예후 및 호흡기능 평가를 위해 사용되며, 환경오염 및 유해물질로 인한 폐의 기능을 판단하기 위해 필요한 검사도구이다(Saputra 등, 2023). 폐활량 검사의 해석은 대상자에게 측정된 데이터와 예측 값을 비교함으로써 이루어진다. 폐활량 예측식은 폐활량 검사 결과를 판정하는 중요한 기준으로 검사결과의 올바른 판단을 위한 정확한 예측 값 마련이 필요하다.

호흡기 질환에 대한 인식이 높아지고 있으나, 호흡기 질환에 노출된 이후 관심을 갖게 되는 문제가 발생하게 된다. 호흡기 질환 특성상 악화와 호전을 반복하며, 지속적인 관리가 필요함에 따라 조기검진이 필요하다(Xu G 등, 2015). 이러한 이유로 ‘제3차(2021~2026년) 국가건강검진 종합 계획’에 의하면 미세먼지 관련 폐질환 조기발견을 위해 일반검진에 폐기능 검사 추가 도입하는 방안을 검토 중이다. 미국흉부학회(American thoracic society)에서는 각 나라별 국민을 대상으로 구한 폐활량 예측식을 목적에 따라 사용할 것을 권고하고 있다. 폐활량 예측식을 선택할 때 인종과 성별, 키, 연령, 생활환경, 측정장비의 정밀성, 평가 프로토콜과 더불어 통계검정 방법의 중요성을 강조하고 있다(Culver 등, 2017; 이재석, 2022).

국내에서의 폐활량 예측식은 중요한 판단 기준임에도 불구하고, 적절한 폐활량 예측식이 없어 동북아시아인 등 외국인을 대상으로 구한 다양한 예측식을 사용하고 있다. 국내에서 사용되는 예측식은 Morris 등(1971), Choi 등(2005), Quanjer 등(2012)이 산출한 예측식을 사용하고 있다. 폐기능 검사기가 대부분 수입제품으로 Morris 등(1971)을 주로 많이 사용한다(대한결핵 및 호흡기학회, 2016). 국내 인구집단을 대상으로 폐활량 검사를 하고 이를 평가하고자 할 때에도 국내 정상인을 대상으로 구한 폐활량 예측식을 사용하는 것이 가장 바람직할 것이다. 이에 Choi 등(2005)에 의해 한국인을 대상으로 폐활량 예측식이 개발되었다. 그러나 약 20년전 환경요인과 생활습관의 변화, 이로 인한 신체조건의 변화와 더불어 연령별 고른 분포가 아닌 주로 20~30대로 여성이 남성보다 약 5배 이상으로 인구통계학적 기준으로 보기에 제한적인 면이 있다.

폐활량 예측치 선정 시 특정 집단을 대상으로 개발되는 등 대상자 선정에서 대표성의 문제와 검사방법이 국제적 기준에 적합하지 않거나, 검사기기의 정확성이 낮고, 검사결과에 대한 재현성이나 신뢰도가 보장되지 않는 경우 등이 지적되었다. 이에 엄격한 대상자 선정과 측정장비의 정밀성, 평가 프로토콜, 그리고 통계검정의 단순회귀식이 아닌 유연한 비선형 모델링을 통한 최적화 가능한 폐활량 예측치가 필요한 실정이다. 따라서, 본 연구는 대전/충청권 지역주민을 대상으로 미국흉부학회에서 제시한 폐활량 검사기와 폐활량 측정방법에 따라 자료의 선정과 정상인의 선정기준을 국제적 기준에 맞게 설정한 폐활량 예측식을 개발하고 새로운 예측모델을 제시하고자 한다.

II. 연구방법

1. 연구대상자

본 연구는 폐활량 예측식 산출에 필요한 새로운 통계검정모델을 제안하기 위한 예비타당도 연구이다. 인구통

계학적 정상 기준을 최대한 반영하기 위하여 2020년 1월부터 22년 7월까지 D광역시 C대학병원에서 폐기능검사를 실시한 지역 20세 이상 성인 남녀 중 정상인 기준에 적합한 데이터를 선별하였다.

선정기준으로 혈액검사, 방사선 검사, 호흡곤란 척도 검사를 통하여 호흡기에 대한 건강여부를 확인하였다. 이후 전문의 문진을 통해 폐기종, 만성기관지염, 천식, 기흉, 결핵, 폐섬유증의 질환판정 유무 및 약물복용 여부, 최근 1개월 이내에 독감 및 COVID-19, 심한 감기 등 급성호흡기질환 판정 유무 및 약물복용 여부등의 병력이 있는 경우는 정상기준에서 제외하였다. 또한 본 연구에 영향을 미칠 수 있는 호흡기 질환을 1개월 이내에 앓고 있거나 감기와 독감, 폐렴 등 호흡기 질환이 있는 자, 호흡기 관련 기관지 확장제 등 약물을 복용중인 자, 협심증, 심근경색 등 심장질환이 있는 자, 현재 흡연중인 자, 경부나 흉부수술의 과거력이 있는 자, 체질량지수(body mass index, BMI) 35 이상인자, 임신한 자는 제외하였다.

검사당일 기침, 객담, 감기, 발열, 피로, 호흡곤란이 있는 경우와 기관지 확장제 등의 약제를 복용하고 있는 경우에 정상인에서 제외하였으나, 과거 흡연력이 있거나 폐활량에 영향을 미칠 정도의 유해인자에 노출된 이력이 있는 경우는 판단기준으로 제시되지 못하였다. 폐기능검사를 실시한 30,058명 중 정상적인 폐기능 소견을 보인 최종 857명의 데이터를 최종 수집하여 통계 분석하였다.

2. 측정방법

폐활량 검사의 정확성과 신뢰할 만한 측정결과를 얻기 위하여 폐기능 검사실에서 근무중인 각각 15년차, 6년차의 임상병리사가 검사를 실시하였다. 폐활량 검사에 사용된 검사기는 국내 결핵 및 호흡기학회에서 제안한 폐활량 검사기로 정확도와 정밀도를 만족하는 V-Max Encore 22(Carefusion, California, USA, 2013)을 사용하였으며, 미국흉부학회/유럽호흡기학회 지침에 따라 검사를 시행하였다(Kreider 등, 2007)

폐활량 검사에 앞서 매 회 3L 실린지로 보정하여 산출된 값이 2.91~3.09 L 이내의 범위로 측정됨을 확인하여 측정값의 정확도를 높였다. 폐활량 검사의 목적과 방법에 대하여 설명하고, 시범을 보여 대상자가 충분히 이해하도록 하였다. 검사 대상자는 의자에 앉은 자세에서 상체를 앞으로 굽히지 않도록 하고, 목을 약간 뒤로 젖힌 상태에서 노이즈 클립으로 코를 막고 검사를 실시하였다. 노력성 호기시 최소한 6초 이상 불어내도록 독려하였고, 마이크로가드 필터와 마우스피스(종이) 필터는 개인마다 교체하여 사용하였다.

미국 흉부학회에서 제시한 폐활량 검사 결과 판단기준에 따라 적합성이 있는 검사를 3번 이상 실시하고, 검사치마다 차이가 200mL 미만으로 재현성 있는 결과 중 가장 큰 검사값을 선택(best value selection method)하였다.

폐활량 검사를 통하여 측정된 지표 중 노력성 폐활량(forced vital capacity, FVC), 1초간 노력성 호기량(Forced Expiratory Volume, FEV1), 1초간 노력성 날숨량비(FEV1/FVC), 그리고 최대호기 유속(Peak expiratory flow, PEF)값에 대한 자료를 수집하였다.

3. 자료처리

통계분석은 R package의 version 4.3.2(R 4.3.2)을 사용하였다. 폐활량 검사지표와 변수들의 상관관계를 파악하여 분포의 특성과 예측치의 방향을 검토한 후 회귀분석을 실시하여 유의한 변수를 선정하였다. 예측인자들의 단위의 상이함으로 인한 회귀계수 추정치의 변동을 줄이고 상호 비교성을 높이기 위하여 각 예측인자를 최소-최대 정규화(min-max normalization)1) 하였다. 일반화가능모형(generalized linear models, GLMs) 적합결과를 토대로 각 예측인자들과 폐활량지표들간의 함수형태를 파악하고 이를 분위수선형회귀의 후보모형 설정에 활용하였다.

성별에 따른 예측인자를 조합하여 각 폐활량지표에서 50분위수 회귀모형을 AIC(Akaike's information Criterion)

분석과 단계별 후방 선택법(stepwise backward selection)을 적용하여 최종모형으로 선택하고, 이를 10분위 모형에도 적용하였다. 후보 모형들에서 AIC값이 가장 낮은 모형을 최종예측식 모형으로 선정하였으며, 성별 예측모형에 포함되는 예측인자를 통일하기 위하여 각 성별에서 선택된 예측인자들을 통합하였다.

Ⅲ. 결 과

1. 연구대상자의 일반적 특성

본 연구에 수집된 대상자의 특성은 폐활량 지표의 대표 예측 인자인 나이, 키, 몸무게와 체질량지수를 고려하여 평균±표준편차로 제시하였다. 연령별 분포는 남성과 여성 모두 60-69세군에서 수집된 데이터가 가장 많았으며 다음 70-79세군, 50-59세군 순으로 데이터가 수집되었다. 분석대상자가 적은 군은 남성에서는 90대 이상군(3명), 30~39세군(12명), 20~29세군(14명) 이었으며, 여성은 20~29세군(10명), 30~39세군(14명)이었다. 남자의 키범위는 144.2cm부터 188.5cm이며 평균 166.9cm이었다. 20~29세군이 가장 컸으며 나이가 증가할수록 감소하였다. 여성의 키범위는 136.2cm에서 177.0cm으로 평균 154.3cm이었으며 남성과 같은 성향을 보였다. 남성의 체중은 평균 66.63kg이었으며 30~39세군에서 가장 높았으며, 나이가 증가할수록 체중은 감소하였다. 여성의 평균체중은 59.28kg으로 40~49세군에서 가장 높았으며 이후 감소하는 성향을 보였다. 체질량지수의 변화는 남성은 30~39세군까지 증가 후 나이가 증가할수록 감소하는 경향을 보였으며, 여성은 50~59세군까지 증가 후 감소하는 경향을 보였다(Table 1).

Table 1. 연령, 그룹 및 성별에 따른 예측인자 지표

Age group	Male(n=552)				Female(n=305)			
	n(%)	Height	Weight	BMI	n(%)	Height	Weight	BMI
20~29	14(2.5)	175.74±6.28	75.99±8.00	24.65±2.81	10(3.2)	161.78±6.26	59.25±9.82	22.67±3.44
30~39	12(2.1)	174.97±4.07	78.31±10.44	25.51±2.63	14(4.5)	160.66±5.45	59.22±14.46	22.63±3.74
40~49	36(6.5)	172.76±3.96	71.53±11.49	24.01±4.00	32(10.4)	160.77±6.03	60.48±10.38	23.09±3.58
50~59	103(18.6)	167.80±6.86	67.07±12.43	23.77±3.93	39(12.7)	157.32±6.49	59.37±10.32	24.04±4.15
60~69	205(37.1)	166.74±5.67	65.80±9.39	23.65±3.03	84(27.5)	154.67±4.92	58.37±8.23	24.44±3.60
70~79	143(25.9)	165.15±6.32	64.89±8.72	23.77±2.75	93(30.4)	149.96±5.17	57.78±8.73	25.71±3.80
80~89	34(6.1)	159.93±7.33	57.11±11.88	22.18±3.41	35(11.4)	150.00±6.37	55.02±10.17	24.31±3.27
≥90~	3(0.5)	155.67±1.22	60.67±6.00	25.06±2.80				

성별과 연령군에 따른 폐활량 지표의 측정값에서 FVC는 남성과 여성 모두 20~29세군에서 가장 높았으며, 나이가 증가할수록 감소하였다. FEV1과 FEV1% 지표에 대해서 남성과 여성 모두 유사한 경향을 보였다. PEF 값은 남성은 20~29세군에서 여성은 40~49세군에서 가장 높았으며 나이가 증가할수록 감소하는 경향을 보였다(Table 2).

Table 2. 연령그룹과 성별에 따른 폐활량측정 지표

Age group	Male(n=552)				Female(n=305)			
	FVC(l)	FEV1	FEV1%	PEF	FVC	FEV1	FEV1%	PEF
20~29	5.07±0.94	3.98±0.61	83.57±2.90	10.32±2.15	3.37±0.35	2.84±0.32	82.00±3.71	6.63±1.08
30~39	4.76±0.23	3.47±0.46	82.25±3.47	10.18±2.18	3.30±0.55	2.73±0.36	80.57±3.55	6.37±1.45
40~49	4.49±0.58	3.34±0.44	77.44±4.85	9.26±1.34	3.24±0.46	2.63±0.40	81.28±4.97	7.57±1.26
50~59	4.12±0.56	3.04±0.20	75.52±4.69	8.89±1.95	2.83±0.57	2.17±0.45	78.21±4.08	5.94±1.44
60~69	3.74±0.67	2.66±0.55	71.07±5.38	8.32±1.99	2.53±0.54	1.96±0.38	77.13±4.57	5.88±1.12
70~79	3.27±0.61	2.27±0.51	69.45±5.93	7.37±2.00	2.26±0.31	1.69±0.33	75.10±4.71	5.35±1.17
80~89	3.11±0.45	2.20±0.33	68.74±5.93	6.71±1.88	2.22±0.44	1.64±0.27	73.21±3.71	4.67±0.90
90~	2.64±0.02	1.89±0.17	68.33±2.52	5.84±2.62				

성별에 따른 예측 인자들과 폐활량 지표들 사이의 상관분석 결과는 Table 3과 같다. 남성과 여성에서 나이는 모든 폐활량지표와 음의 상관을 보였으며, 키와 체중은 FEV1%을 제외한 폐활량지표에서 모두 양의 상관을 보였다. 체질량지수는 상관관계의 규칙성을 나타내지 않았다. 예측인자들 중 몸무게와 체질량지수의 상관성이 매우 크고($r>0.8$) 남자에서는 키와 몸무게의 상관성 또한 강하게 나타났다($r>0.5$). 이와 같은 예측인자들 사이의 강한 상관성으로 인한 다중공선성(multicollinearity) 문제를 고려하여 예측식을 선정하는 과정에서 몸무게와 체질량지수가 동시에 후보 변수로 추천되는 경우 체질량지수를 우선적 선택하였다.

Table 3. 성별에 따른 나이, 키, 몸무게, 체질량지수와 폐활량지표들의 표본상관계수

	Male(n=552)				Female(n=305)			
	Age	Height	Weight	BMI	Age	Height	Weight	BMI
Age								
Height	-0.46*				-0.59*			
Weight	-0.32*	0.55*			-0.16*	0.38*		
BMI	-0.11*	0.06	0.87*		0.19*	-0.18*	0.83*	
FVC	-0.56*	0.49*	0.35*	0.13	-0.55*	0.58*	0.22*	-0.12
FEV1	-0.59*	0.44*	0.26*	0.05	-0.62*	0.56*	0.17*	-0.16
FEV1%	-0.18*	0.01	-0.11*	-0.14*	-0.30*	0.09	-0.09	-0.15*
PEF	-0.40*	0.37*	0.34*	0.19*	-0.46*	0.41*	0.17*	-0.07

예측인자들의 단위의 상이함으로 인한 회귀계수 추정치의 변동을 줄이고 상호 비교성을 높이기 위하여 각 예측인자를 단위에 상관없이 0~1사이에 위치할 수 있도록 최소-최대 정규화(min-max normalization)1) 하였다. 국가 통계포털에서 제공하는 관련 자료들의 분포를 참고하여 아래와 같이 설정하였다.

① 나이 최소-최대 척도화(통계청, 2020)

$$age' = \frac{age - age_{\min}}{age_{\max} - age_{\min}} = \frac{age - 18}{100 - 18}$$

② 키 최소-최대 척도화(국가통계포털, 2020)

$$\text{남성} : height' = \frac{height - height_{\min}}{height_{\max} - height_{\min}} = \frac{height - 140}{200 - 140}$$

$$\text{여성} : height' = \frac{height - height_{\min}}{height_{\max} - height_{\min}} = \frac{height - 130}{185 - 130}$$

③ 몸무게 최소-최대 척도화(국가통계포털, 2020)

$$\text{남성} : weight' = \frac{weight - weight_{\min}}{weight_{\max} - weight_{\min}} = \frac{weight - 35}{120 - 35}$$

$$\text{여성} : weight' = \frac{weight - weight_{\min}}{weight_{\max} - weight_{\min}} = \frac{weight - 30}{95 - 30}$$

④ 체질량지수 최소-최대 척도화(국가통계포털, 2013)

$$BMI = \frac{BMI - BMI_{\min}}{BMI_{\max} - BMI_{\min}} = \frac{BMI - 13}{35 - 13}$$

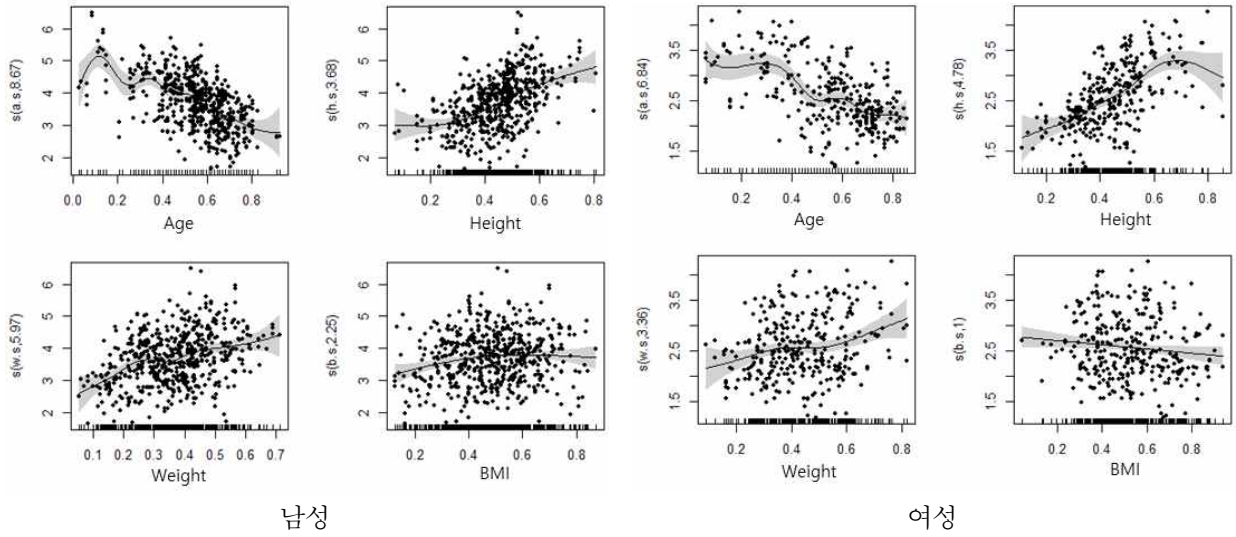
2. FVC 예측모형

예측식에 포함되는 예측인자들의 다항차수를 선택하기 위하여 폐활량 지표와 각 예측인자에 일반화가법모형 (generalized additive models)을 적합시켰다(Tibshirani, 2017)(Figure 3).

연령과 FVC예측 값과의 관계는 20-30대에서 증가패턴을 보이다 연령이 증감함에 따라 점차 감소하는 경향을 보였다. 키와 체중의 증가에 따라 FVC가 증가하고 있었으며, 남자의 경우 BMI가 오름에 따라 상승하는 경향을 보였지만 여성의 경우 낮아지는 경향을 나타냈다. 이는 연령의 증가에 따라 폐활량이 감소하는 연령효과(aging effect)와 체격효과(body-physical effect)를 복합적으로 나타내는 모델이다.

각 예측인자의 일반화가법모형에서 곡선의 형태가 나타나고 있어 본 연구에서는 예측인자들의 다항차수로 최대 2차항까지 포함하였다. 예측모형의 선정은 남자와 여자에서 각각 진행하였다. 본 연구에서 예측하는 모형은 폐활량 정상군에서의 분위수 회귀모형(quantile regression models)으로 예를 들면, 정상군의 하위 10% 또는 50% 수준 등에 대한 예측모형을 추론하는 것이 목적이다(Chernozhukov 등, 2022).

Figure 3. 성별에 따른 FVC와 나이, 키, 몸무게, 체질량지수와 일반화가법모형 적합결과



FVC와 유의한 상관관계가 있는 성별에 따른 예측 인자들을 세 가지 - (1) 나이, 키, 몸무게, (2) 나이, 키, 체질량지수, (3) 나이, 키, 몸무게, 체질량지수 군으로 구성하였다. 각 군에서 아카이케 정보기준(Akaike Information Criterion, AIC)을 사용하여 단계별 후방 선택법(stepwise backward selection)으로 각 군의 후보모형을 선정하였다. 예측인자를 조합하였을 때 AIC값은 남자의 경우 나이, 키, 체질량지수를 포함한 모델의 값이 가장 작았고 여성에서는 연령, 키, 체중, BMI를 포함한 모델의 값이 가장 작아 최종모형으로 선정하였다(Table 4).

Table 4. FVC 예측인자 구성과 AIC분석 결과

성별	예측인자 구성	FVC 50-분위수 회귀모형	AIC 값
남성	Age, Height, Weight	$a + a^2 + h + h^2 + w + w^2$	1097.8
		$a^2 + h + h^2 + w + w^2$	1095.8
		$a^2 + h^2 + w + w^2$	1094.2
	Age, Height, BMI	$a + a^2 + h + h^2 + b + b^2$	1088.0
		$a^2 + h + h^2 + b + b^2$	1086.1
		$a^2 + h^2 + b + b^2$	1084.3
	Age, Height, Weight, BMI	$a + a^2 + h + h^2 + w + w^2 + b + b^2$	1090.6
		$a^2 + h + h^2 + w + w^2 + b + b^2$	1088.6
		$a^2 + h + h^2 + w + w^2 + b$	1086.6
여성	Age, Height, Weight	$a + a^2 + h + h^2 + w + w^2$	425.4
		$a + a^2 + h + h^2 + w$	423.4
		$a + h + h^2 + w$	421.9
		$a + h + w$	420.7
	Age, Height, BMI	$a + a^2 + h + h^2 + b + b^2$	419.3
		$a + h + h^2 + b + b^2$	417.3
		$a + h + b + b^2$	415.4
	Age, Height, Weight, BMI	$a + a^2 + h + h^2 + w + w^2 + b + b^2$	420.3
		$a + a^2 + h + h^2 + w + w^2 + b^2$	418.3
		$a + h + h^2 + w + w^2 + b^2$	416.3
		$a + h + w + w^2 + b^2$	414.4
		$a + h + w + b^2$	412.8

a: age, h: height, w: weight, b: BMI

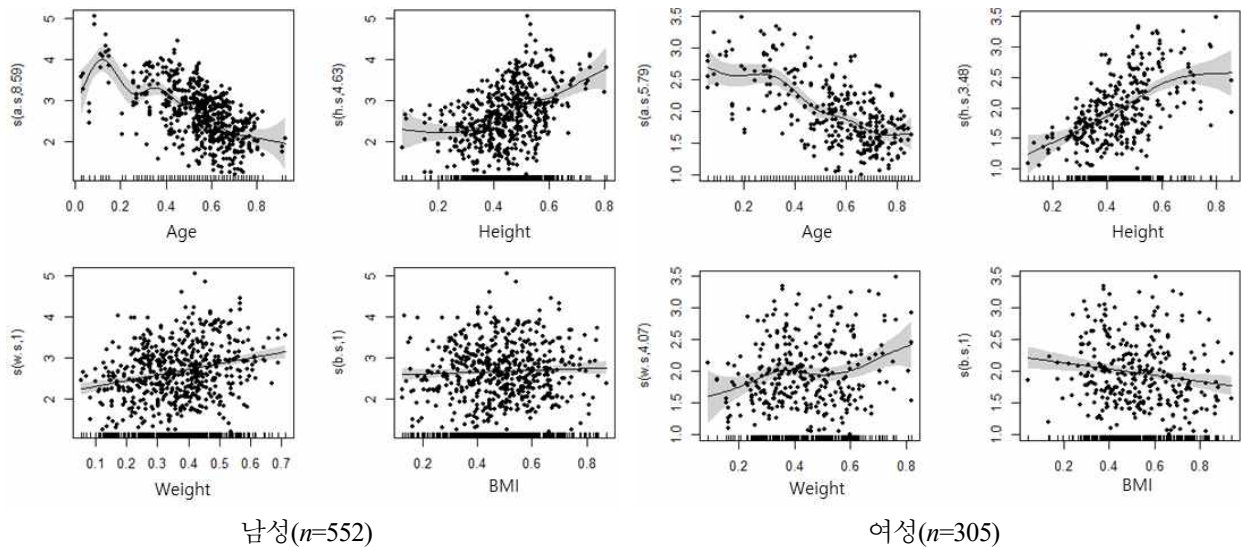
FVC 지표에 대한 분위수에 따른 최종 회귀 추정식은 다음과 같다.

성별	50-분위수회귀추정식($\tau=0.5$)
남성	$\widehat{Q}_\tau(y) = \widehat{\beta}_0(\tau) + \widehat{\beta}_1(\tau)*a + \widehat{\beta}_2(\tau)*a^2 + \widehat{\beta}_3(\tau)*h + \widehat{\beta}_4(\tau)*h^2 + \widehat{\beta}_5(\tau)*b + \widehat{\beta}_6(\tau)*b^2$ $= 2.964 - 0.019*a - 2.104*a^2 - 0.348*h + 2.673*h^2 + 4.547*b - 4.332*b^2$
여성	$\widehat{Q}_\tau(y) = \widehat{\beta}_0(\tau) + \widehat{\beta}_1(\tau)*a + \widehat{\beta}_2(\tau)*a^2 + \widehat{\beta}_3(\tau)*h + \widehat{\beta}_4(\tau)*h^2 + \widehat{\beta}_5(\tau)*b + \widehat{\beta}_6(\tau)*b^2$ $= 2.964 - 0.019*a - 2.104*a^2 - 0.348*h + 2.673*h^2 + 4.547*b - 4.332*b^2$
성별	10-분위수회귀추정식($\tau=0.1$)
남성	$\widehat{Q}_\tau(y) = \widehat{\beta}_0(\tau) + \widehat{\beta}_1(\tau)*a + \widehat{\beta}_2(\tau)*a^2 + \widehat{\beta}_3(\tau)*h + \widehat{\beta}_4(\tau)*h^2 + \widehat{\beta}_5(\tau)*b + \widehat{\beta}_6(\tau)*b^2$ $= 1.664 + 1.009*a - 2.908*a^2 - 1.604*h + 3.999*h^2 + 6.117*b - 5.864*b^2$
여성	$\widehat{Q}_\tau(y) = \widehat{\beta}_0(\tau) + \widehat{\beta}_1(\tau)*a + \widehat{\beta}_2(\tau)*a^2 + \widehat{\beta}_3(\tau)*h + \widehat{\beta}_4(\tau)*h^2 + \widehat{\beta}_5(\tau)*b + \widehat{\beta}_6(\tau)*b^2$ $= 2.321 - 4.282*a + 3.583*a^2 + 1.642*h - 0.349*h^2 + 0.339*b - 0.059*b^2$

3. FEV₁ 예측모형

성별에 따른 FVC와 예측인자들간 일반화가법모형(GLMs) 적합결과 남성과 여성 모두 30세 이후 연령의 증가에 따라 감소하는 경향을 보였으며, 키와 체중이 증가함에 따라 상승하였고, BMI의 증가에 따라 남성은 증가, 여성은 감소하는 경향을 나타내었다(Figure 5).

Figure 5. 성별에 따른 FEV₁와 나이, 키, 몸무게, 체질량지수와 일반화가법모형 적합결과



그래프 형태로 볼 때 각 예측인자의 일반화가법모형에서 곡선의 형태가 나타나고 있어 마찬가지로 예측인자

들의 다항차수로 최대 2차항까지 포함하였다. FEV1과 유의한 상관관계가 있는 성별에 따른 예측인자 구성은 남녀 모두 나이, 키, 몸무게였으며 AIC값이 가장 낮은 예측인자 구성은 남성은 나이, 키, 체질량지수, 여성은 나이, 키, 몸무게, 체질량지수 군이었다(Table 5)

Table 5. FEV1 예측인자 구성과 AIC분석 결과

성별	공변량 구성	FEV1 50-분위수 회귀모형	AIC 값
남성	Age, Height, Weight	$a + a^2 + h + h^2 + w + w^2$	924.3
		$a + a^2 + h^2 + w + w^2$	923.3
		$a + a^2 + h^2 + w^2$	921.9
	Age, Height, BMI	$a + a^2 + h + h^2 + b + b^2$	918.4
		$a + a^2 + h^2 + b + b^2$	917.3
	Age, Height, Weight, BMI	$a + a^2 + h + h^2 + w + w^2 + b + b^2$	918.9
		$a + a^2 + h + h^2 + w + w^2 + b$	918.0
		$a^2 + h + h^2 + w + w^2 + b$	917.4
여성	Age, Height, Weight	$a + a^2 + h + h^2 + w + w^2$	318.2
		$a + a^2 + h + h^2 + w$	317.1
		$a + a^2 + h + h^2$	315.2
	Age, Height, BMI	$a + a^2 + h + h^2 + b + b^2$	316.1
	Age, Height, Weight, BMI	$a + a^2 + h + h^2 + w + w^2 + b + b^2$	319.8
		$a + a^2 + h + h^2 + w + w^2 + b$	317.8
		$a + a^2 + h^2 + w + w^2 + b$	316.0

성별에서 선택된 예측인자들을 통합하여 이를 최종모형의 예측인자로 하여 추정된 FEV1에 대한 분위수회귀 모형은 아래와 같다.

$$Q_\tau(y) = \beta_0(\tau) + \beta_1(\tau) \times a + \beta_2(\tau) \times a^2 + \beta_3(\tau) \times h^2 + \beta_4(\tau) \times b + \beta_5(\tau) \times b^2$$

FEV1 지표에 대한 분위수에 따른 최종 회귀추정식은 다음과 같다.

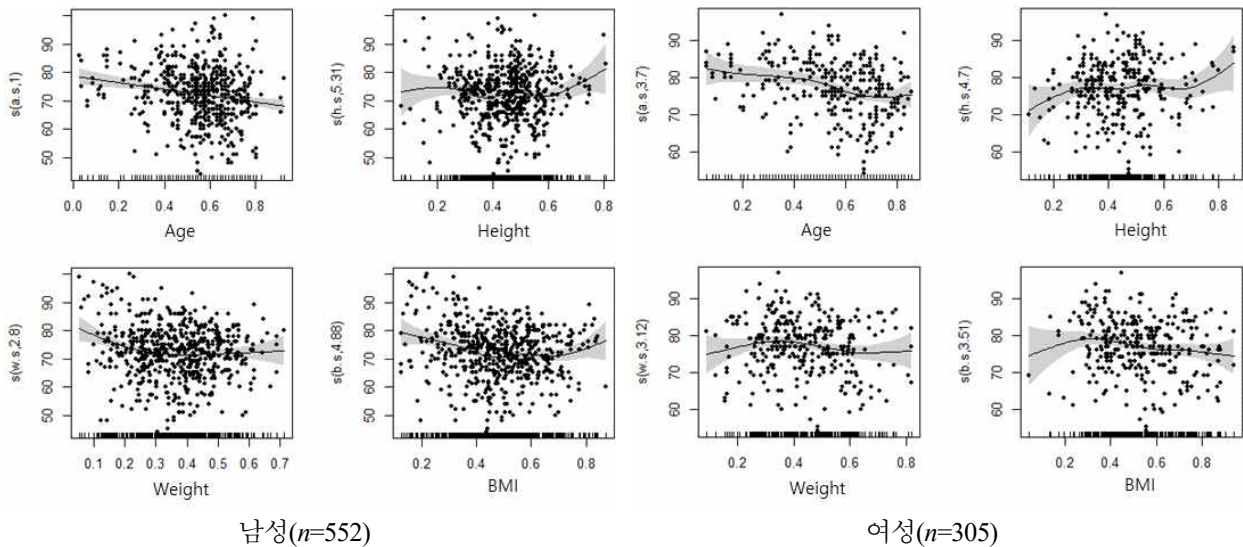
성별	50-분위수회귀추정식($\tau=0.5$)
남성	$\widehat{Q}_\tau(y) = \widehat{\beta}_0(\tau) + \widehat{\beta}_1(\tau) * a + \widehat{\beta}_2(\tau) * a^2 + \widehat{\beta}_3(\tau) * h^2 + \widehat{\beta}_4(\tau) * b + \widehat{\beta}_5(\tau) * b^2$ $= 2.792 - 0.809 * a - 1.401 * a^2 + 1.510 * h^2 + 2.399 * b - 2.584 * b^2$
여성	$\widehat{Q}_\tau(y) = \widehat{\beta}_0(\tau) + \widehat{\beta}_1(\tau) * a + \widehat{\beta}_2(\tau) * a^2 + \widehat{\beta}_3(\tau) * h^2 + \widehat{\beta}_4(\tau) * b + \widehat{\beta}_5(\tau) * b^2$ $= 1.885 - 1.756 * a + 0.639 * a^2 + 1.454 * h^2 + 1.986 * b - 1.662 * b^2$

성별	10-분위수회귀추정식($\tau=0.1$)
남성	$\widehat{Q}_\tau(y) = \widehat{\beta}_0(\tau) + \widehat{\beta}_1(\tau)*a + \widehat{\beta}_2(\tau)*a^2 + \widehat{\beta}_3(\tau)*h^2 + \widehat{\beta}_4(\tau)*b + \widehat{\beta}_5(\tau)*b^2$ $= 1.747 - 0.489*a - 1.045*a^2 + 2.117*h^2 + 1.479*b - 1.301*b^2$
여성	$\widehat{Q}_\tau(y) = \widehat{\beta}_0(\tau) + \widehat{\beta}_1(\tau)*a + \widehat{\beta}_2(\tau)*a^2 + \widehat{\beta}_3(\tau)*h^2 + \widehat{\beta}_4(\tau)*b + \widehat{\beta}_5(\tau)*b^2$ $= 2.132 - 3.126*a + 2.193*a^2 + 0.804*h^2 + 0.852*b - 0.942*b^2$

4. FEV₁% 예측모형

성별에 따른 FEV₁%와 예측인자들간 일반화가법모형(GLMs) 적합결과 남성과 여성 모두 연령의 증가에 따라 감소하는 경향을 보였으며, 키와 체중이 증가함에 따라 상승하였고, BMI의 증가에 따라 남성은 증가, 여성은 감소하는 경향을 나타내었다(Figure 7). 그래프 형태로 볼 때 각 예측인자의 일반화가법모형에서 곡선의 형태가 나타나고 있어 마찬가지로 예측인자들의 다항차수로 최대 2차항까지 포함하였다.

Figure 7. 성별에 따른 FEV₁%와 나이, 키, 몸무게, 체질량지수와 일반화가법모형 적합결과



FEV₁%와 유의한 상관관계가 있는 성별에 따른 예측인자 구성은 남성은 나이, 체중, 체질량지수였으며, 여성은 나이와 체질량지수였다. 예측인자를 조합하였을 때 AIC값이 가장 낮은 모형은 남성의 경우 나이, 키, 체중, 체질량지수 군이었으며, 여성은 나이, 키, 몸무게군이었다 (Table 6).

Table 6. FEV1% 예측인자 구성과 AIC분석 결과

성별	공변량 구성	FEV1% 50-분위수 회귀모형	AIC 값
남성	Age, Height, Weight	$a + a^2 + h + h^2 + w + w^2$	4043.1
		$a^2 + h + h^2 + w + w^2$	4041.3
		$a^2 + h + w + w^2$	4040.4
		$a^2 + w + w^2$	4040.1
	Age, Height, BMI	$a + a^2 + h + h^2 + b + b^2$	4038.1
		$a + h + h^2 + b + b^2$	4036.1
	Age, Height, Weight, BMI	$a + a^2 + h + h^2 + w + w^2 + b + b^2$	4034.8
		$a + h + h^2 + w + w^2 + b + b^2$	4032.9
여성	Age, Height, Weight	$a + a^2 + h + h^2 + w + w^2$	2097.4
		$a + a^2 + h^2 + w + w^2$	2095.5
		$a^2 + h^2 + w + w^2$	2093.6
		$a^2 + h^2 + w^2$	2092.3
	Age, Height, BMI	$a + a^2 + h + h^2 + b + b^2$	2098.3
		$a + a^2 + h^2 + b + b^2$	2096.4
		$a^2 + h^2 + b + b^2$	2094.4
		$a^2 + h^2 + b^2$	2092.6
	Age, Height, Weight, BMI	$a + a^2 + h + h^2 + w + w^2 + b + b^2$	2101.0
		$a + a^2 + h + h^2 + w + b + b^2$	2099.0
		$a^2 + h + h^2 + w + b + b^2$	2097.1
		$a^2 + h + w + b + b^2$	2095.2
		$a^2 + w + b + b^2$	2093.8
		$a^2 + w + b$	2093.4

선택된 예측인자를 통합하여 최종모형의 예측인자로 설정 후 성별에 따른 FEV1%예측식을 분위수 회귀모형으로 추정하였다. 추정된 분위수회귀모형은 다음과 같다.

$$Q_{\tau}(y) = \beta_0(\tau) + \beta_1(\tau) \times a + \beta_2(\tau) \times a^2 + \beta_3(\tau) \times h + \beta_4(\tau) \times h^2 + \beta_5(\tau) \times w + \beta_6(\tau) \times b$$

FEV1% 지표에 대한 분위수에 따른 회귀추정식은 다음과 같다.

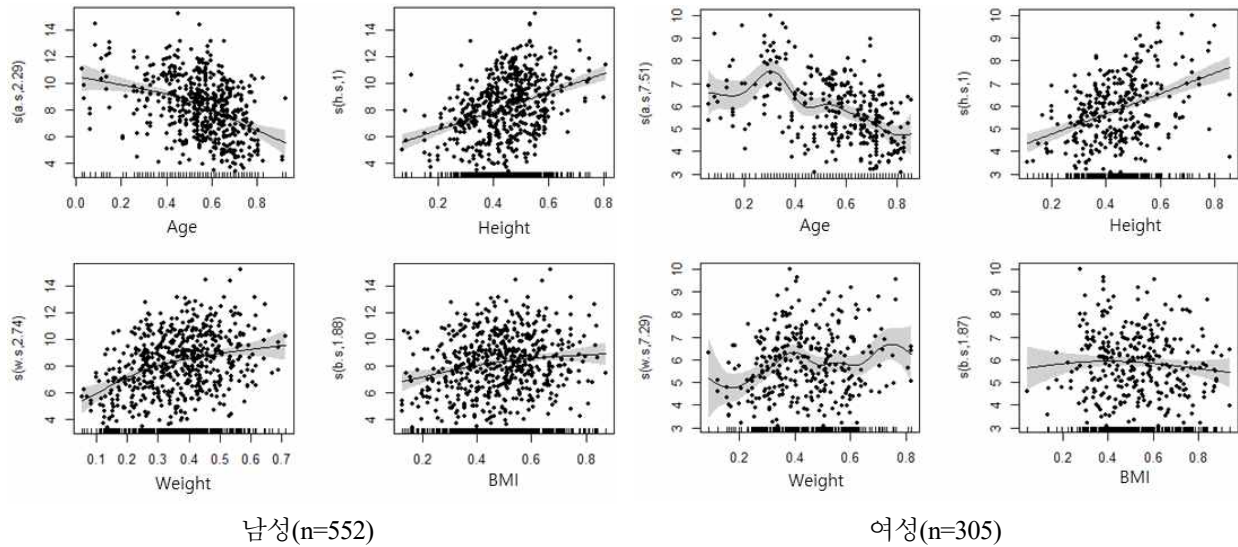
성별	50-분위수회귀추정식($\tau=0.5$)
남성	$\hat{Q}_\tau(y) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)*a + \hat{\beta}_2(\tau)*a^2 + \hat{\beta}_3(\tau)*h + \hat{\beta}_4(\tau)*h^2 + \hat{\beta}_5(\tau)*w^2 + \hat{\beta}_6(\tau)*b$ $= 99.700 - 9.629*a - 5.441*a^2 - 52.077*h + 39.358*h^2 + 16.854*w^2 - 14.564*b$
여성	$\hat{Q}_\tau(y) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)*a + \hat{\beta}_2(\tau)*a^2 + \hat{\beta}_3(\tau)*h + \hat{\beta}_4(\tau)*h^2 + \hat{\beta}_5(\tau)*w^2 + \hat{\beta}_6(\tau)*b$ $= 81.093 - 7.314*a - 6.897*a^2 + 15.459*h - 18.033*h^2 - 16.116*w^2 + 6.313*b$

성별	10-분위수회귀추정식($\tau=0.1$)
남성	$\hat{Q}_\tau(y) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)*a + \hat{\beta}_2(\tau)*a^2 + \hat{\beta}_3(\tau)*h + \hat{\beta}_4(\tau)*h^2 + \hat{\beta}_5(\tau)*w^2 + \hat{\beta}_6(\tau)*b$ $= 96.408 - 1.624*a - 22.189*a^2 - 61.338*h + 39.239*h^2 + 48.972*w^2 - 34.211*b$
여성	$\hat{Q}_\tau(y) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)*a + \hat{\beta}_2(\tau)*a^2 + \hat{\beta}_3(\tau)*h + \hat{\beta}_4(\tau)*h^2 + \hat{\beta}_5(\tau)*w^2 + \hat{\beta}_6(\tau)*b$ $= 58.631 - 39.398*a + 20.450*a^2 + 53.545*h - 28.526*h^2 - 65.335*w^2 + 40.064*b$

5. PEF 예측모형

성별에 따른 PEF와 예측인자들간 일반화가법모형(GLMs) 적합 결과 남성은 연령의 증가에 따라 감소하는 경향을 보였으며, 여성은 40대까지 증가후 감소하는 경향을 보였다. 남성과 여성 모두 키와 체중 그리고 체질량지수 증가에 따라 상승하는 경향이 나타났다(Figure 8). 그래프 형태로 볼 때 각 예측인자의 일반화가법모형에서 곡선의 형태가 나타나고 있어 마찬가지로 예측인자들의 다항차수로 최대 2차항까지 포함하였다.

Figure 8. 성별에 따른 PEF와 나이, 키, 몸무게, 체질량지수와 일반화가법모형 적합결과



PEF와 유의한 상관관계가 있는 예측인자 구성은 남성은 나이, 키, 체중, 체질량지수였으며, 여성은 나이, 키, 체중이었다. 예측인자를 조합하였을 때 AIC값이 가장 낮은 모형은 남성의 경우 나이, 키, 체중 군이었으며, 여성은 나이, 키, 체중, 체질량지수 군이었다(Table 7).

Table 7. PEF 예측인자 구성과 AIC분석 결과

성별	공변량 구성	PEF 50-분위수 회귀모형*)	AIC 값	
남성	Age, Height, Weight	$a + a^2 + h + h^2 + w + w^2$	2360.9	
		$a + a^2 + h^2 + w + w^2$	2359.0	
		$a^2 + h^2 + w + w^2$	2358.5	
	Age, Height, BMI	$a + a^2 + h + h^2 + b + b^2$	2365.1	
		$a + a^2 + h + b + b^2$	2363.1	
		$a^2 + h + b + b^2$	2362.8	
	Age, Height, Weight, BMI	$a + a^2 + h + h^2 + w + w^2 + b + b^2$	2364.1	
		$a + a^2 + h + w + w^2 + b + b^2$	2362.2	
		$a + a^2 + h + w + b + b^2$	2360.2	
		$a^2 + h + w + b + b^2$	2359.4	
	여성	Age, Height, Weight	$a + a^2 + h + h^2 + w + w^2$	1032.2
			$a + a^2 + h + w + w^2$	1030.3
Age, Height, BMI		$a + a^2 + h + h^2 + b + b^2$	1033.8	
		$a + a^2 + h + b + b^2$	1032.0	
Age, Height, Weight, BMI		$a + a^2 + h + h^2 + w + w^2 + b + b^2$	1034.7	
		$a + a^2 + h + h^2 + w + w^2 + b$	1032.8	
		$a + a^2 + h + w + w^2 + b$	1031.0	
		$a + a^2 + w + w^2 + b$	1029.5	

선택된 예측인자를 통합하여 최종모형의 예측인자로 설정 후 성별에 따른 PEF 예측식을 추정한 분위수 회귀 모형은 다음과 같다.

$$Q_{\tau}(y) = \beta_0(\tau) + \beta_1(\tau) \times a + \beta_2(\tau) \times a^2 + \beta_3(\tau) \times h^2 + \beta_4(\tau) \times w^2 + \beta_5(\tau) \times b$$

PEF 지표에 대한 분위수에 따른 회귀추정식은 다음과 같다.

성별	50-분위수회귀추정식($\tau=0.5$)
남성	$\hat{Q}_{\tau}(y) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau) * a + \hat{\beta}_2(\tau) * a^2 + \hat{\beta}_3(\tau) * h^2 + \hat{\beta}_4(\tau) * w^2 + \hat{\beta}_5(\tau) * b$ $= 4.663 + 5.051 * a - 9.979 * a^2 + 8.845 * h^2 - 10.491 * w^2 + 7.717 * b$
여성	$\hat{Q}_{\tau}(y) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau) * a + \hat{\beta}_2(\tau) * a^2 + \hat{\beta}_3(\tau) * h^2 + \hat{\beta}_4(\tau) * w^2 + \hat{\beta}_5(\tau) * b$ $= 3.946 + 4.382 * a - 7.457 * a^2 + 4.948 * h^2 - 4.513 * w^2 + 3.787 * b$
성별	10-분위수회귀추정식($\tau=0.1$)
남성	$\hat{Q}_{\tau}(y) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau) * a + \hat{\beta}_2(\tau) * a^2 + \hat{\beta}_3(\tau) * h^2 + \hat{\beta}_4(\tau) * w^2 + \hat{\beta}_5(\tau) * b$ $= 2.946 + 4.910 * a - 7.906 * a^2 + 6.380 * h^2 - 1.894 * w^2 + 2.999 * b$
여성	$\hat{Q}_{\tau}(y) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau) * a + \hat{\beta}_2(\tau) * a^2 + \hat{\beta}_3(\tau) * h^2 + \hat{\beta}_4(\tau) * w^2 + \hat{\beta}_5(\tau) * b$ $= 5.203 - 2.183 * a - 0.180 * a^2 + 0.848 * h^2 + 0.475 * w^2 + 0.216 * b$

IV. 논 의

본 연구는 한국인 폐활량 예측산식을 위한 예비타당성 연구로써, 정상 폐기능 소견을 지닌 대전/충청권 지역의 20세 이상 성인 남녀를 대상으로 수집한 폐활량 검사지표를 성별에 따른 예측인자들과 조합하여 통계검정 모델 기반의 예측식 산출방법을 제안하기 위함이다.

폐활량 예측식은 대상자의 인구통계학적 요인, 폐활량 검사 절차와 방법, 검사기기, 예측인자 및 통계검정 방법 등에 따라 연구자마다 결과에 대한 차이가 있다. 동일 인구집단을 대상으로 산출된 예측값일지라도 차이가 있을 수 있기에 미국 흉부학회의 폐활량 검사기준과 역학적 기준, 통계학적 기준을 고려해야 한다. 본 연구에서는 이러한 조건들을 충족하기 위하여 미국 흉부학회에서 제시한 표준화된 검사기기를 사용하여 검사 절차를 수행하였으며, 대상자의 폐기능 정상여부를 판단하기 위하여 흉부 X-ray와 문진을 통해 폐에 대한 건강정보를 확인하였다.

수집된 30,058명의 데이터 중 정상적인 폐기능 소견을 보인 데이터는 857명으로 약 2.85%로 집계되었다. 이중 남자는 552명으로 64.41%, 여자는 305명으로 35.58%이었다. 수집된 데이터의 연령대별 분포는 20대 2.80%, 30대 3.03%, 40대 7.93%, 50대 17.03%, 60대 34.07%, 70대 28.12%, 80대 8.16%, 90대 0.35%으로 나타났으며 60대와 70대의 연령층에서 수집된 데이터가 전체 데이터의 50%이상을 차지하였다. 국내에서 사용 중인 대표적인 폐활량예측식의 선행연구인 Choi 등(2005)에서는 20-30대의 데이터 분포가 수집된 데이터의 50%이상을 보이는 것과는 달리 비교적 높은 연령대의 데이터를 중심으로 예측식이 마련된 것을 확인할 수 있다.

대표적인 폐활량 지표 예측인자로는 성별, 연령, 신장이 있으며, 신체조건의 다양화와 급변화로 인해 점차 체중도 상관성이 있는 인자로 반영되고 있다. Morris (1971)의 연구에서는 연령과 키 인자만으로 예측식을 산출하였으며, Choi 등(2005)의 연구에서는 연령의 2승, 신장, 체중 인자를 포함하여 예측식을 산출하였다. 본 연구결과에서도 체중은 폐활량 지표와 유의한 상관이 있었으며, 이에 더하여 신체조건의 다양화를 고려한 체질량지수 또한 일부 폐활량 지표와 상관성이 있음을 확인하였다. 하지만 체중과 체질량지수와의 강한 상관성으로 인해 통계모델의 안정성을 해치는 다중공선성(multicollinearity) 문제를 고려하지 않을 수 없었다. 다중공선성이 높을 경우 회귀계수의 추정치가 불안정해지고 통계검정력이 감소될 수 있다는 여러 선행연구의 결과를 참고하여 예측식을 선정하는 과정에서 몸무게와 체질량지수가 동시 변수로 제시되는 경우 체질량 지수를 우선적으로 선택하였다(Belsley 등, 2005). 또한 예측인자들의 단위가 상이함으로 이로 인한 회귀계수 추정치의 변동성을 줄이고 상호 비교성을 높이기 위하여 각 예측인자를 단위에 상관없이 0~1사이에 위치할 수 있도록 최소-최대 정규화(min-max normalization) 과정을 통해 입력하였다(Hastie 등, 2009). 최소-최대정규화(min-max normalization) 방법은 모든 데이터가 같은 정도의 스케일로 반영되도록 해주는 것을 목적으로 하며 데이터들을 균일하게 만드는 가장 일반적인 방법이다. 데이터의 분포는 유지하면서 비교적 일정한 범위를 정규화하는데 사용된다(Loukas, 2020).

일반화가법모형Generalized Linear Models (GLMs)은 다양한 종류의 응답변수에 대한 예측을 수행하는데 주로 사용되는 통계 모델 중 하나로 선형 회귀 모델을 확장하여 비선형성 및 분포의 다양성을 다루기 위해 고안되었다(Dobson과 Barnett, 2018; McCullagh와 Nelder, 1989). 본 연구에서는 일반화가법모형 적합결과를 토대로 각 예측인자들과 폐활량지표들간의 함수형태를 파악하고 분위수선형회귀의 후보모형 설정에 활용하였다. 분위수 선형 회귀(Quantile Regression)는 종속 변수의 특정 분위수에 대한 조건부 중앙값 또는 분위수를 예측하는 통계적 모델링 방법이다(Koenker, 2005). 종속변수의 평균에 대한 예측을 수행하는 일반적인 선형회귀와는 달리 분위수 선형 회귀는 종속변수의 특정 분위수에 대한 예측을 수행함으로써 중앙값(50%) 예측 뿐만 아니라, 다른 분위수에 대한

예측이 가능하다. 또한, 이상치에 민감한 선형회귀모델에 비해 상위, 하위 극단값의 영향을 받지 않으며, 종속 변수의 분포에 대한 가정이나 정규성 가정이 없어 데이터의 분포에 대해 더 유연하게 대응할 수 있다. 이에 본 연구에서는 50분위수 회귀모형으로 정상폐활량지표의 중앙예측값을 제시하였으며, 상 하위 10% 범주 이내 예측식 또한 제시하였다. 표준화된 예측식을 제공하는 것 이외에도 검사결과를 해석하는 방법은 지속적으로 논쟁 중이다. 임상학회에서는 정상과 비정상 검사 결과를 95%의 참조 구간을 기준으로 하는 반면, 호흡기학회에서는 측정값을 예측값의 %으로 표현하는 것이 오랜 관습이 되어오고 있다. 이러한 전통은 Bates와 Christie(1964)의 권고에서 비롯되었다. 일반적으로 유용한 규칙은 예측된 정상값에서 예측값 주변의 분산이 표준편차가 10%이내에 근접하고 평균예측값에 비례하는 경우에 유효할 것이라고 제안하였다.(Miller, 2007; Miller와 Pincock, 1988; Miller 등, 2011). 수집된 데이터를 일반화 가법모형 결과로 제시한 성별에 따른 연령별 폐활량 지표 간 그래프에서 FVC, FEV1 모두 20대 중반까지 상승한 이후 고령화에 따라 점차 감소하는 패턴을 보였으며, FEV1%와 PEF는 고령화 될수록 지속적으로 감소하는 패턴을 보였다. 이러한 통계검정 모델의 결과는 Global lung function initiative 2012(GLI-2012) 네트워크에서 소개된 인종별 연령에 따른 폐활량 지표 패턴과 유사하였다. GLI-2012는 33개국 72개 센터를 통해 수집한 1600만개 이상의 폐활량 지표 중 97,759개의 기록을 토대로 인종별 연령(3~95세)에 따른 폐활량 예측식을 제시하였으며 이는 국제호흡기학회의 승인을 받아 많은 국가 내 권장 참조 예측식으로 선정되었다. 하지만 한국은 중국(북쪽)과 함께 북아시아인으로 포함되어 수집된 4,992개의 데이터를 기반으로 예측식이 제시되고 있다.

선행 예측식들과 본 연구를 통해 산출된 예측식 간의 잔차 분석은 아래와 같은 공식을 사용하여 잔차율(%)으로 비교하였다.

$$\text{Residue}(\%) = 100 * \{(\text{선행예측값} - \text{산출예측값}) / \text{산출예측값}\}$$

그 결과 남녀 모두 선행연구를 통해 산출된 예측값의 잔차율이 모두 양의 관계를 보이고 있어 본 예측식이 상대적으로 낮은 예측치를 추정하고 있음을 확인할 수 있다. 남성의 FVC의 경우 최정근식과 비교하였을 때 20-30대의 잔차율이 10% 미만이나 40대 이후 10% 이상의 차이를 나타내며 고연령대로 갈수록 잔차율은 더욱 커졌다. Morris식과의 비교에서는 20-30대에서 가장 컸으며 고령으로 갈수록 잔차 비율이 줄어드는 경향을 보였다. FEV1의 경우 최정근 등의 식은 고연령대로 갈수록 점차 높아지다 낮아지는 경향을 보였으며, Morris식과의 비교에서는 고연령대로 갈수록 잔차율이 줄어드는 경향을 보였다. 여성의 FVC와 FEV1도 비슷한 경향을 보였으나 Morris식과 비교하여 60대 이상 고연령대에서는 음의 관계를 보이기도 하였다(Table 8).

Table 8. 선행 예측식과의 잔차 분석 결과

Age group	Male (n=552)				Female (n=305)			
	FVC		FEV ₁		FVC		FEV ₁	
	Choi(%)	Morris(%)	Choi(%)	Morris(%)	Choi(%)	Morris(%)	Choi(%)	Morris(%)
20~29	7.49	9.92	15.74	15.98	10.77	14.16	13.64	15.29
30~39	8.67	7.19	16.59	13.53	12.82	13.02	17.09	15.72
40~49	10.16	6.46	17.23	11.81	13.21	9.46	18.77	14.19
50~59	10.38	2.99	15.58	8.50	13.71	5.25	18.40	10.88
60~69	12.74	1.85	14.92	6.48	12.87	-0.03	15.19	5.26
70~79	16.31	1.99	14.00	5.62	11.98	-7.52	6.15	-6.52
80~89	19.78	2.79	10.15	4.26	9.03	-13.99	-5.00	-18.09
90~	27.20	3.59	4.52	5.83				

Choi 등(2005)에 의해 국내 건강 성인을 대상으로 예측산식이 제안되었고 그들의 예측식 또한 과거 선행연구 및 해외에서 제안한 아시아인 예측식 보다 높게 추정된 바 있다. 한국인의 체중과 신장과 같은 인구통계학적 지표가 상향 조정된 요인이 예측값을 높게 산정하는게 기인했을 것이라는 것이 연구자의 생각이다. 하지만 그들의 예측값은 고연령대로 갈수록 더욱 높게 추정되는 경향이 있어 고령층 수준의 폐기능을 호흡기장애로 오인하는 의학적 진단이 내려질 수 있다.

Chou 등(2005)의 연구 이후 약 20년의 시간이 흐르면서 대한민국 성인의 신장과 체중 조건이 높아진 것은 사실이다. 하지만 본 연구에서 수집된 대상자의 평균 신장과 체중은 선행연구의 대상자들보다 낮게 측정되었는데 그 이유는 수집된 연령대별 가중치의 차이에 기인하는 것으로 여겨진다. Choi 등(2005)의 연구에서는 남녀 모두 20-30대에서 수집된 데이터가 50%이상을 차지하고 있는 반면 본 연구는 60-70대에서 수집된 데이터가 50%이상을 차지하고 있어 고령화 연령대에 데이터가 치우쳐 있음을 알 수 있다. 이렇듯 인구통계학적 데이터 수집이 특정 연령대로 편향되어 평균신장과 체중의 차이가 발생되었을 뿐 아니라 폐활량지표의 예측 또한 낮게 산출된 것으로 사료된다. 예측값이 낮게 산출된 두 번째 이유는 대기 환경오염과 COVID-19과 같은 호흡기질환 확산으로 인한 유해환경 노출대상자 선별 문제에 있다. Zhao 등(2020)의 연구에서는 COVID-19 감염 이후 55명의 환자를 추적관찰한 결과 25.45%인 14명에게서 폐기능의 이상을 보고하였으며, Huang 등(2020)의 연구에서는 57명의 COVID-19의 환자를 대상으로 폐기능검사를 실시한 결과 75.4%에 해당하는 43명에게서 폐기능 이상을 확인하였다. COVID-19를 경험했던 대부분의 대상자들이 폐활량의 저하를 호소하였다고 한다. 이는 본 연구의 데이터 수집과정에서 선정조건 이외에 분류하지 못한 제외기준을 웃도는 COVID-19의 과거력과 직업환경적 요소의 고려가 충분히 고려되지 못했기 때문이라고 생각한다. 세번째 이유는 수집된 데이터 분포의 대표성 문제이다. 본 연구의 데이터는 대전/충청 지역 대상자들만으로 수집되었기에 국내 전역의 건강한 성인을 대표한다고 가정하기 어렵다. 이에 따라 권역별 인구수 대비 연령대별 건강한 성인의 폐활량지표를 기반으로 한 인구통계학적 데이터가 수집 될 필요가 있다.

이미 전 세계적으로 400개가 넘는 예측방정식이 발표되었다. 방정식 산출에 필요한 건강한 피험자 선정기준, 측정 장비, 검사 프로토콜, 품질관리 그리고 통계적 접근 방식과 같은 요인으로 인해 적합한 방정식 채택은 어려

움이 있다. 폐기능 예측식 산식의 오랜 문제 중 하나는 어린시절부터 노년기까지 환자를 원활하게 모니터링 할 수 있는 단일 소스가 부족하다는 점이다. 전 연령대에 걸친 모집단을 모집하는 것이 어려웠고, 생애주기 동안 변하는 폐 크기, 연령, 신장, 체중 간의 관계를 고려하여 데이터를 모델링 하는 것이 어려웠다. 특히 가장 큰 장애는 사용가능한 통계적 방법론이 제약적이었던 것이다. 수십년 동안 제안된 폐활량 예측식은 폐활량지표와 연령, 키 사이의 관계를 설명하기 위해 단순선형회귀 기법을 사용하여 참조 방정식을 도출했기 때문에 동일한 방정식을 사용하여 어린이와 성인 모두 폐기능을 예측하기는 어려웠다. 하지만 최근 보다 유연한 방법론이 등장하면서 다양한 연령대에 걸쳐 복잡한 비선형관계를 모델링 할 수 있게 되었다.

빠른 고령화로 인해 호흡기질환과 같은 노인성 질환의 유병률은 높아지고 있다(Marcus 등, 2016; Miner 등, 2016). 호흡기 증상이 있는 고령자의 경우 호흡기장애의 근본적인 메카니즘을 파악하기 위해서는 높은 진단 정확도가 필수적이다(Boyd 등, 2005; Fired 등, 2012; Marcus 등, 2016; Miner 등, 2016; Vaz Fragoso 등, 2013). 고령층에 맞는 정상적인 폐기능을 호흡기장애로 오인하는 것을 최소화함으로써(Vaz Fragoso 등, 2015), 고령자에게 부적절하고 잠재적으로 유해한 호흡기 약물의 사용을 피할 수 있을 뿐만 아니라 호흡기 증상이 있는 고령자에서 다른 진단을 고려하는데 지연이 발생하는 것을 예방할 수 있다(Boyd 등, 2005; Fired 등, 2012; Marcus 등, 2016; Miner 등, 2016; Vaz Fragoso 등, 2013). 2012년 GLI보고서(Quanjer 등, 2012)에 따르면, 진단 정확도를 더욱 높이기 위해서는 참조집단 중 70대 연령의 고령자의 대표성을 높여야 한다. 80세 이상에서는 생존편향의 문제가 있어 일반화를 위한 정보가 제한적이기 때문이다. 고령자의 폐기능을 해석할때는 주의가 필요하며, 검사수행의 적절성과 관련하여 연령에 적합한 기준을 개발하고 폐기능의 대체 측정치를 평가하기 위해 고령자 대상의 더 많은 데이터가 필요하다. 이러한 개선을 통해 노인성 질환에서 호흡기질환 검사 결과의 일반화 가능성을 넓힐 수 있다.

V. 결 론

본 연구는 한국인 폐활량 예측산식을 위한 예비타당성 연구로 정상 폐기능 소견을 지닌 지역사회 20세 이상 성인 남녀를 대상으로 폐활량 검사지표를 통계검정 모델기반의 예측산식을 제시하였다. 미국흉부학회에서 제시하고 있는 연령과 신장변수에 체중이 포함되어 차이가 있었다. 본 예측식에서도 체중은 폐활량 지표와 유의한 상관관이 있으며, 이에 더하여 신체조건의 다양화를 고려한 체질량지수 또한 폐활량 지표와 상관성을 확인하였다. 본 연구의 분위수 선형 회귀는 50분위수 회귀모형으로 정상폐활량지표의 중앙예측값을 제시하였으며, 상하위 10% 범주 이내 예측식을 제시하였다. 선행 예측식 간의 잔차 비율은 모두 양의 관계를 보이고 있으며, 본 예측식은 고령화 연령대의 데이터, COVID-19와 같은 환경적 요인 등으로 상대적으로 낮은 예측치를 추정하고 있음을 시사한다. 이는 호흡기 증상이 있는 고령자의 진단 정확도를 높이기 위한 임상적 의의가 있는 예비타당성 연구라 할 수 있다.

이해 충돌

본 연구의 저자들은 연구, 저작권 및 출판과 관련하여 잠재적인 이해충돌이 없음을 선언합니다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1C1C101350111).

참고문헌

- 국가통계포털. 한국인 인체지수 조사 성인 몸무게. 국가통계포털. 2020
- 국가통계포털. 한국인 인체지수 조사 성인 키. 국가통계포털. 2020
- 국가통계포털. 한국인 인체지수 조사 체질량지수. 국가통계포털. 2013
- 대한결핵 및 호흡기학회. 폐기능검사지침. 대한결핵 및 호흡기학회 2016.
- 최정근, 백도명, 이정오. 한국인의 정상 폐활량 예측치. 대한결핵 및 호흡기학회지 2005;58(3):230-42.
- 통계청. 한국인의 인구구조 현황 2020.
- 이재석, 강태욱. 딥벨트를 이용한 일회성 가로막 호흡운동이 가로막 움직임과 노력성 폐활량에 미치는 영향. 대한물리치료과학회지.2022;29(2):57-65.
- Bates DV, Christie RV. "Respiratory Function in Disease: An Introduction to the Integrated Study of the Lung: With the assistance of Margaret R. Becklake (and others)." 1964;144.
- Belsley, David A, Edwin Kuh, et al. Welsch. Regression diagnostics: Identifying influential data and sources of collinearity. John Wiley & Sons, 2005.
- Boyd CM, Darer J, Boult C, et al. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. JAMA 2005;294:716-724.
- Chernozhukov V, Fernández-Val I, Melly B. Fast algorithms for the quantile regression process. Empirical Economics. 2022;1-27.
- Culver BH, Graham BL, Coates AL, et al. Recommendations for a standardized pulmonary function report. An official American Thoracic Society technical statement. Am J Respir Crit Care Med 2017;196(11):1463-72.
- Dobson, Annette J, Adrian GB. An introduction to generalized linear models. CRC press, 2018.
- Fried TR, Vaz Fragoso CA, Rabow MW. Caring for the older person with chronic obstructive pulmonary disease. JAMA 2012;308:1254-1263.
- Hastie T, Tibshirani R, Friedman JH et al. The elements of statistical learning: data mining, inference, and prediction. New York: springer, 2009.

- Koenker, R. "Quantile regression Cambridge University Press New York." 2005.
- Kreider, M.-E.; Grippi, M.-A. Impact of the new ATS/ERS pulmonary function test interpretation guidelines. *Respir. Med.* 2007, 101, 2336-2342.
- Marcus BS, McAvay G, Gill TM, et al. Respiratory symptoms, spirometric respiratory impairment, and respiratory disease in middle-aged and older persons. *J Am Geriatr Soc* 2015;63:251-257.
- McCullagh, P, Nelder JA. "Generalized linear models CRC press." Boca Raton. USA: Florida, 1989.
- Miller MR. What defines abnormal lung function? *Thorax* 2007;62:1107.
- Miller MR, Pincock AC. Predicted values: how should we use them? *Thorax* 1988;43:265-267.
- Miller MR, Quanjer PH, Swanney MP, et al. Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2011;139:52-59.
- Miner B, Tinetti ME, Van Ness PH, et al. Dyspnea in community-dwelling older persons: a multifactorial geriatric health condition. *J Am Geriatr Soc* 2016;64:2042-2050.
- Morris JF, Koski A, Johnson LC. Spirometric standards for healthy nonsmoking adults. *Am Rev Respir Dis* 1971;103(1):57-67.
- Quanjer PH, Stanojevic S, Cole TJ, et al Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. 2012;1324-43.
- Ranu H, Wilde M, Madden B. "Pulmonary function tests". *Ulster Med J* 2011;84.
- Saputra PB, Purwati DD, Ulhaq AUD, et al Neutrophil Elastase in the Pathogenesis of Chronic Obstructive Pulmonary Disease: A Review. *Current Respir Med Rev* 2023;19(1):29-35.
- Tibshirani. Robert John. Generalized additive models. Routledge 2017.
- Vaz Fragoso CA, Gill TM, McAvay G, et al. Respiratory impairment in older persons: when less means more. *Am J Med* 2013;126:49-57.
- Vaz Fragoso CA, McAvay G, Van Ness PH, et al. Phenotype of normal spirometry in an aging population. *Am J Respir Crit Care Med* 2015;192:817-825.
- Xu G, Chen Z, Cao X, et al Analysis of pulmonary function test results in a health check-up population. *J Thorac Dis* 2015;7(9):1624.
- Y Huang, C Tan, J Wu, et al. Impact of coronavirus disease 2019 on pulmonary function in early convalescence phase. *Respi res* 2020;21:1-10.
- Y Zhao, Y Shang, W song, et al. Follow-up study of the pulmonary function and related physiological characteristics of COVID-19 survivors three months after recovery. *EClinicalMedicine* 2020;25.
-