Original Article

Goodness-of-Fits of the Spirometric Reference Values for Koreans and USA Caucasians to Spirometry Data from Residents of a Region within Chungbuk Province

Sang Yong Eom, Ph.D.[1], Sun In Moon, M.S.[1], Dong Hyuk Yim, M.S.[1], Chul Ho Lee, Ph.D.[2], Guen Bae Kim, Ph.D.[3], Yong Dae Kim, Ph.D.[1], Jong Won Kang, M.D., Ph.D.[1], Kang Hyeon Choe, M.D., Ph.D.[4], Sung Jin Kim, M.D., Ph.D.[5], Byung Sun Choi, M.D., Ph.D.[6], Seung Do Yu, Ph.D.[3], Soung Hoon Chang, M.D., Ph.D.[7], Jung Duck Park, M.D., Ph.D.[6], Heon Kim, M.D., Ph.D.[1]

[1]Department of Preventive Medicine and Medical Research Institute, Chungbuk National University College of Medicine, Cheongju, [2]Asbestos Damage Relief Center, Korea Environment Coporation, [3]Environmental Health Research Division, Environmental Health Research Department, National Institute of Environmental Research, Incheon, Departments of [4]Internal Medicine and [5]Radiology, Medical Research Institute, Chungbuk National University College of Medicine, Cheongju, [6]Department of Preventive Medicine, Chung-Ang University College of Medicine, Seoul, [7]Department of Preventive Medicine, Konkuk University College of Medicine, Chungju, Korea

Background: Korean regression models for spirometric reference values are different from those of Americans. Using spirometry results of Korean adults, goodness-of-fits of the Korean and the USA Caucasian regression models for forced vital capacity (FVC) and forced expiratory volume in one second (FEV₁) were compared.

Methods: The number of study participants was 2,360 (1,124 males and 1,236 females). Spirometry was performed under the guidelines of the American Thoracic Society and the European Respiratory Society. After excluding unsuitable participants, spirometric data for 729 individuals (105 males and 624 females) was included in the statistical analysis. The estimated FVC and FEV₁ values were compared with those measured. Goodness-of-fits for Korean and USA Caucasian models were compared using an F-test.

Results: In males, the expected values of FVC and FEV₁ using the Korean model were 12.5% and 5.7% greater than those measured, respectively. The corresponding values for the USA Caucasian model were 3.5% and 0.6%. In females, the difference in FVC and FEV₁ were 13.5% and 7.7% for the Korean model, and 6.3% and 0.4% for the USA model, respectively. Goodness-of-fit for the Korean model regarding FVC was not good to the study population, but the Korean regression model for FEV₁, and the USA Caucasian models for FVC and FEV₁ showed good fits to the measured data.

Conclusion: These results suggest that the USA Caucasian model correlates better to the measured data than the Korean model. Using reference values derived from the Korean model can lead to an overestimation regarding the prevalence of abnormal lung function.

Key Words: Spirometry; Reference Values; Vital Capacity; Forced Expiratory Volume

Introduction

Spirometry test is a very important tool in assessing overall pulmonary health status and a clinical test used widely for diagnosis and clinical evaluation of respiratory diseases, epidemiological investigations and clinical researches[1]. In recent days, the use of spirometry is gradually increasing in epidemiological investigations on health problems related with inhalation of environmental harmful materials.

Address for correspondence: **Heon Kim, M.D., Ph.D.**
Department of Preventive Medicine and Medical Research Institute, Chungbuk National University College of Medicine, 52, Naesudong-ro, Heungdeok-gu, Cheongju 361-763, Korea
Phone: 82-43-261-2864, Fax: 82-43-274-2965
E-mail: kimheon@chungbuk.ac.kr

Most countries have their own estimation models for the prediction of pulmonary function indexes, and the predictive values are used as the diagnostic criteria for abnormal lung function. USA has induced various regression models for pulmonary function parameters according to different ethnic groups and gender[2], which has been widely used not only in USA but in Korea as well. In Korea, there have been a few studies inducing predictive regression models for pulmonary function parameters in general population[3,4]. Choi et al.[4] had presented predictive models which were induced with spirometry data from 4,816 subjects of the 2001 Korean National Health and Nutrition Examination Survey. Different variables were used in the Korean and the American regression models for forced vital capacity (FVC) and forced expiratory volume in one second (FEV$_1$). Choi et al.[4] had used age, height and body weight as independent variables in the models for FVC and FEV$_1$. In contrast, Hankinson et al.[2] had age and height only, excluding the body weight from the variables. Predictive values were also different. The values for Americans were greater than those for Koreans in younger adults under 40 years of age, but smaller in

40 years old or older individuals. In over 50 years of age group, the gap between the two predictive values gets wider as age increases. Therefore, which model is used as the reference for spirometry test of individuals older than 40 years of age is a major determinant of the prevalence of ventilatory impairment.

The USA regression models are now being widely used as the references for spirometry test results of Koreans. This fact leads to a necessity for examination of goodness-of-fits of the Korean and the USA regression models to spirometric data of healthy Korean adults. The purpose of this study is to compare the goodness-of-fits of the Korean and USA regression models to spirometry data of Korean subjects who lived in Chungbuk Province.

Materials and Methods

1. Study subjects

From July 1, 2010 to August 11, 2010, the authors conducted an environmental epidemiological investigation on the health effects of dust and gas emitted from cement plants located in Jecheon-si and
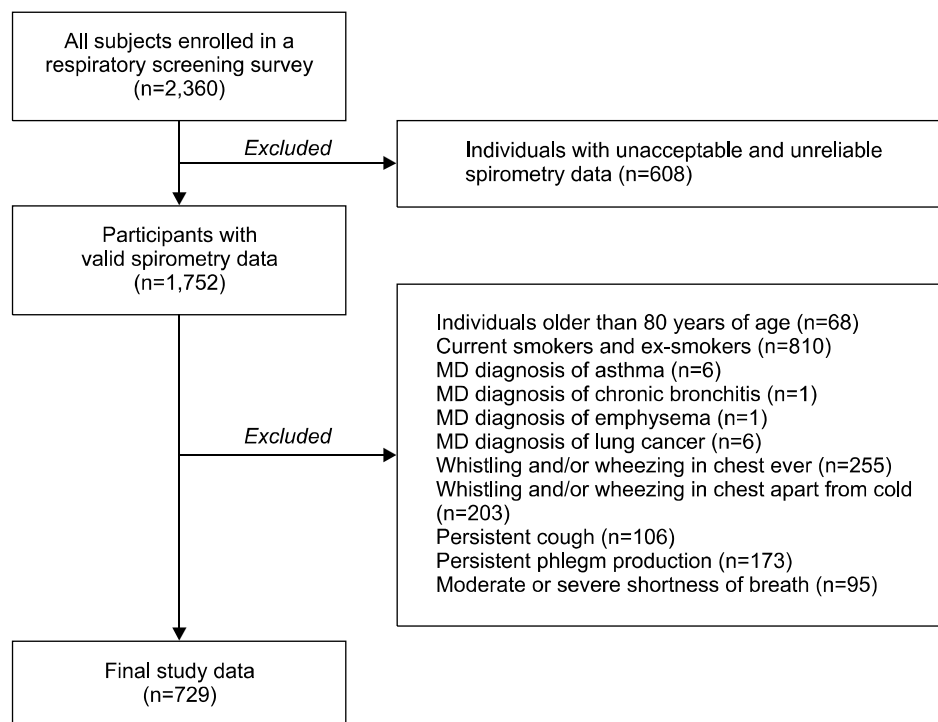

Figure 1. Flow chart with details concerning the selection of study subjects. MD: medical doctor.

Danyang-gun, Chungbuk Province, Republic of Korea. As a part of this investigation, the respiratory function tests were performed to subjects who were 40 years old or older, and lived in the vicinity of the cement plants.

A questionnaire was implemented to gather information on demographic characteristics, smoking and drinking habits as well as respiratory symptoms, which was followed by the respiratory examination in the order of spirometry test, chest radiography and physical examination. Voluntarily, 2,360 adults (1,124 males and 1,236 females) were participated in this examination. After the investigation was completed, 6 experts determined on the validity of the spirometry tests and then excluded 608 individuals whose spirometric tests were not valid. As the result, 1,752 individuals had moved unto the next stage (Figure 1).

2. Investigation methods

1) Questionnaire survey and measurement of heights and body weights:
Using a structured questionnaire, information was gathered on the demographic factors, smoking habit, occupational history, past medical history, and presence/absence of respiratory and cardiac diseases at the time of test. Thereafter, height and body weight of the subjects were measured. For height measurement, every subject was asked to put off the shoes, mount on an extensometer gathering heels and stand to adhere his/her back to the vertical axis of extensometer. Because the investigation was performed in summer, the subject's weight was measured as they wore light casual clothes without outers. Based on the results of questionnaire survey, those applicable to any of the followings (Table 1) were excluded.

Finally, data of 729 individuals (105 males, 624 females) was included in the statistical analysis (Figure 1).

2) Chest radiography:
The thoracic radiography was taken using a digital radiography system (Digi SPOT DR-XG; Sitech Medical, Virginia Gardens, FL, USA), and the images were read by two specialized radiologists. When there was any discrepancy between the radiologist's diagnoses, another radiologist performed additional reading for final determination.

Table 1. Numbers of subjects excluded using exclusion criteria

Individuals 40 years of age and older (n=1,752)	Number excluded
Individuals older than 80 years of age	68
Current smokers and ex-smokers	810
MD diagnosis of asthma	6
MD diagnosis of chronic bronchitis	1
MD diagnosis of emphysema	1
MD diagnosis of lung cancer	6
Whistling and/or wheezing in chest ever	255
Whistling and/or wheezing in chest apart from cold	203
Persistent cough	106
Persistent phlegm production	173
Moderate or severe shortness of breath	95
Number remained	729

MD: medical doctor.

3) Spirometry test:
Spirometry test was performed using Micro Quark PC-based Spirometers (Cosmed, Rome, Italy) according to the guidelines of American Thoracic Society and European Respiratory Society for accuracy and reliability[1]. In the spirometry test, FVC, FEV_1, and peak expiratory flow rate (PEFR) were measured. All the spirometers were calibrated using a 3 L calibration syringe every morning before starting examination and every 4 hours. The air temperature and humidity of the testing venue were measured and the values were entered into the laptop system connected to the spirometers.

Before testing, an explanation on spirometry test was given to the subjects. Spirometry tests were repeated until the reproducibility and goodness of the test reached the criteria of American Thoracic Society and European Respiratory Society. First, there should be no coughing during the 1st second of expiration, no glottic obstruction influencing measurement results, no early completion, no early cut-off, no air leak and no mouthpiece blocking. Secondly, the subject did his/her best without hesitation in starting expiration, and the extrapolated volume at the beginning of expiration should be less than 5% of FVC or 0.15 L. Thirdly, the subjects should exhale over 6 seconds or until the volume-time curve had reached to the plateau or no more expiration

was possible. When the discrepancy between the best value and the second best value among best 3 FVC measurements were within 0.15 L, and, at the same time, the discrepancy between the best value and the second best value among best 3 FEV$_1$ measurements was within 0.15 L, the spirometry test result was determined as valid. All subjects had been tested ≥ 3 times at least, and nearly all subjects were tested more than 5 times. Since the flow-volume curve and the flow-time curve were displayed on the monitor while spirometry test was in-progress, testers and testees could check expiration time. At each time of test completion, the computer determined the validity of the test, and all valid test results were numerically saved in the computer. When the best FEV$_1$/FVC% was less than 70%, re-test was performed after inhalation of a bronchodilator according to the decision of the examining physician.

After completion of spirometry test, the flow-volume curve and the flow-time curve, the test results for FVC, FEV$_1$, PEFR and the reference value of National Health and Nutrition Examination Survey (NHANES III) were printed out and submitted to the examining physician, who determined if any additional tests or inhalation of bronchodilator would be required. In this study, 3 to 4 testers conducted spirometry test at the same time using spirometers of the same model, and less than 40 subjects were allocated to every tester per day.

The best FVC and the FEV$_1$ were selected respectively from 3 or more accurate and reproducible tests. After the spirometry tests were completed for all subjects, 6 experts in spirometry test had determined whether the each test result was compatible to the accuracy and reproducibility criteria of American Thoracic Society.

4) Statistical analysis: Predictive values of FVC and FEV$_1$ were calculated by applying parameters such as age, height and weight of the subjects to the Korean regression models and to the USA Caucasian regression models, and compared to actually measured FVC and FEV$_1$. A paired t-test was used for statistical comparison between those mean values. Scatter diagrams for the FVC and FEV$_1$ values measured, those estimated using

the Korean regression models and those estimated using the USA Caucasian regression models according to age or height changes were drawn for males and females, respectively. Trend lines for measured values, those for estimated values using the Korean regression models and those for estimated values using the USA Caucasian regression models were added to the scatter diagrams. Sum of squares of differences between the measured value and the predictive values of the two regression models (sum of square due to error [SSE]) and sum of squares of differences between the mean of measured values and the estimated values of the two regression models (total sum of square [SST]) were obtained respectively and the goodness-of-fits for those models were statistically tested. To test validities of the regression models, we used the following equation including SSE and sum of square due to regression (SSR) which equals to SST minus SSE, and an F-test.

$$F = \frac{\dfrac{SSR}{\text{Degree of freedom of SSR}}}{\dfrac{SSE}{\text{Degree of freedom of SSE}}}$$

F-values obtained by the following equation were utilized to compare the goodness-of-fits of the two regression models.

$$F = \frac{\dfrac{\text{SSE of the Korean regression model}}{\text{Degree of freedom of SSE of the Korean regression model}}}{\dfrac{\text{SSE of the USA Caucasian regression model}}{\text{Degree of freedom of SSE of the USA regression model}}}$$

Results

Means and standard deviations of age, height, weight, and measured and estimated FVC and FEV$_1$ values of the study subjects are presented in Table 2. The means of estimated values derived from the Korean models overestimated more than those from the USA models on all occasions.

Scatter diagrams and trend lines for the FVC and FEV$_1$ values measured, those estimated using the Korean re-

Table 2. Means and standard deviations of age, height, weight, FVC, and FEV_1 in subjects

			Mean	Standard deviation	Ratio to measured value	Test for difference to measured value (p-value)
Males (n=105)	Age		59.1	10.6		
	Height		165.5	5.7		
	Weight		66.4	8.5		
	FVC	Measured	3.848	0.736		
		Estimated using the Korean model*	4.328	0.409	1.125	<0.001
		Estimated using the USA Caucasian model[†]	3.983	0.553	1.035	<0.001
	FEV_1	Measured	3.005	0.644		
		Estimated using the Korean model	3.265	0.467	1.057	<0.01
		Estimated using the USA Caucasian model	3.024	0.506	1.006	NS
Females (n=624)	Age		60.1	9.5		
	Height		152.2	5.7		
	Weight		58.0	8.9		
	FVC	Measured	2.590	0.498		
		Estimated using the Korean model	2.940	0.348	1.135	<0.001
		Estimated using the USA Caucasian model	2.789	0.432	1.063	<0.001
	FEV_1	Measured	2.156	0.470		
		Estimated using the Korean model	2.292	0.353	1.077	<0.001
		Estimated using the USA Caucasian model	2.164	0.381	1.004	NS

*Choi et al.[4]. [†]Hankinson et al.[2].
FVC: forced vital capacity; FEV_1: forced expiratory volume in one second.

gression models and those estimated using the USA Caucasian regression models according to age or height changes were shown in Figure 2 (males) and Figure 3 (females).

Since the SSE of the Korean regression model for FVC in the male subjects of this present study was greater than its SST, the Korean regression model for FVC turned out to be not valid for males (Table 3). For FEV_1, whereas, the goodness-of-fits of the Korean male model and the USA male model were statistically significant (p <0.01).

The SSE of the Korean model for FVC in males was significantly greater than that of the corresponding USA model. For FEV_1, the SSEs of the Korean and the American models did not show any statistical difference (Table 3). In females, the USA regression models were valid for FVC and FEV_1, but the Korean model was not for FVC. The SSEs of the USA models for FVC and FEV_1 were significantly smaller than the corresponding Korean models.

Discussion

The FVC and FEV_1 values of the Korean regression models for the male subjects of this study were higher than the measured values by 12.5% for FVC and by 3.5% for FEV_1, respectively. Whereas, the values estimated with the USA Caucasian regression models were higher only by 5.7% for FVC and by 0.6% for FEV_1. This means that FVC and FEV_1 estimated using the USA Caucasian models were closer to the measured values than those using the Korean models. In particular, the means of FEV_1 measured and FEV_1 estimated using the USA male models were almost same. Female data showed a similar pattern, but sizes of overestimation of the Korean models were larger than in males. The difference between the mean of measured FEV_1 and that of FEV_1 estimated using the American female model was only 0.4%.

Because of the larger difference between the measured FVC value and the FVC estimated with the Korean
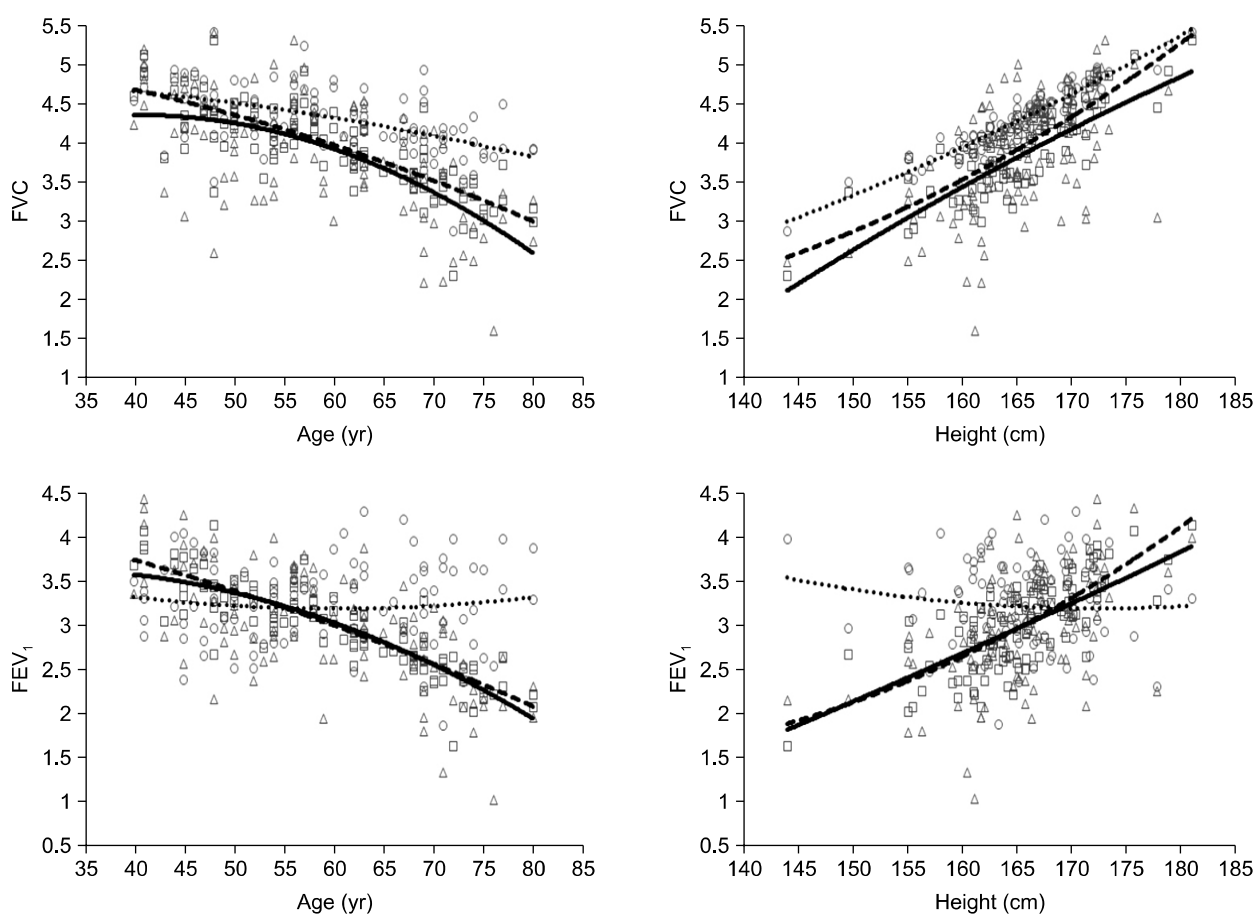
Figure 2. Scatter diagrams for FVC and FEV$_1$ values measured ($\triangle$), those estimated using the Korean regression model ($\bigcirc$) and estimated using the USA Caucasian regression model ($\square$) in male subjects. Trend line — is for measured values, ⋯⋯ for estimated values using the Korean regression models, and ---- for estimated values using the USA Caucasian regression models, respectively. FVC: forced vital capacity; FEV$_1$: forced expiratory volume in one second.

regression model, if the Korean model is used as a reference, then more subjects would be classified as restrictive ventilatory impairment. Moreover, the larger difference between the measured and estimated FVC values than that between those FEV$_1$ values lead to an increase in the number of subjects to be classified as obstructive ventilatory impairment. Those over-diagnoses would be more prominent in females than in males.

The regression curves for the American model were closer to the curves for measured values than those for the Korean models not only in the FVC and FEV$_1$ scattergrams including age in the X axis but also in those graphs including height as the X variable. The regression curves for measured FEV$_1$ and for FEV$_1$ esti-

mated with the USA Caucasian regression model almost overlapped each other. Taking into account the minimal difference between the mean values of measured FEV$_1$ and estimated FEV$_1$ using the USA models, these USA models would be more appropriate as the reference for Koreans than the Korean models.

Because the SSR of Korean model for FVC was nearly equal to the SSE, the R-square of the model is actually 0. This means that the Korean models for FVC do not fit to measured FVC data of this present study at all. In case of FEV$_1$, the Korean regression models showed significant goodness-of-fits. The goodness-of-fits of the USA regression models were statistically significant for both FVC and FEV$_1$ in both males and females, and the R-squares of the models were higher than correspond-
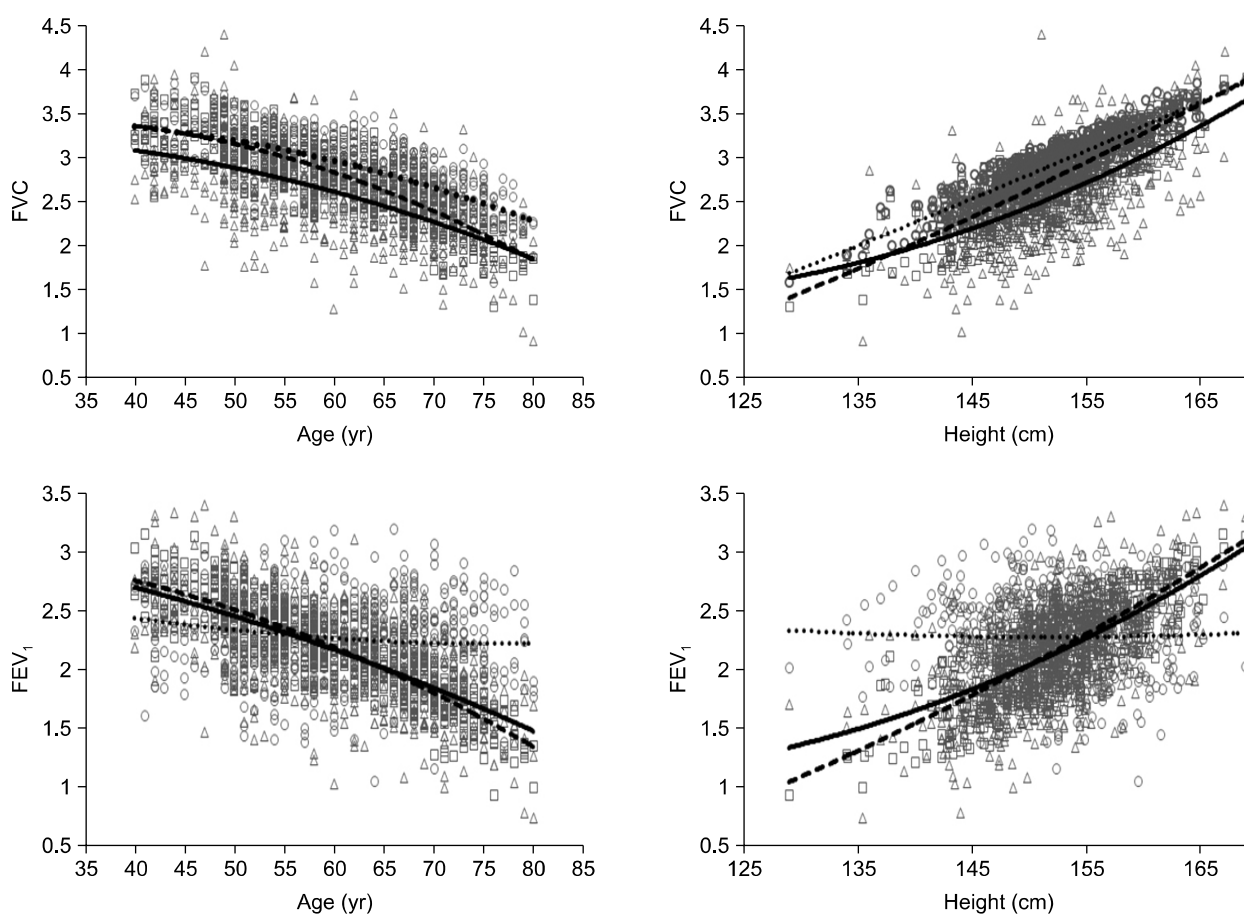
Figure 3. Scatter diagrams for FVC and FEV_1 values measured ($\triangle$), estimated using the Korean regression model ($\bigcirc$) and estimated using the USA Caucasian regression model ($\square$) in female subjects. Trend line $-$ is for measured values, $\cdots\cdots$ for estimated values using the Korean regression models, and ---- for estimated values using the USA Caucasian regression models, respectively. FVC: forced vital capacity; FEV_1: forced expiratory volume in one second.

Table 3. SSTs, SSEs, and goodness-of-fits of Korean regression models and USA Caucasian regression models for FVC and FEV_1

			SST	SSE	p-value for the fit to measured value	p-value for comparison between the two models
Males (n=105)	FVC	Measured	56.38			
		Korean model*		57.30	Not valid	<0.001
		USA model[†]		27.41	<0.01	
	FEV_1	Measured	43.15			
		Korean model		27.38	<0.01	NS
		USA model		20.41	<0.01	
Females (n=624)	FVC	Measured	154.40			
		Korean model		154.00	NS	<0.001
		USA model		101.93	<0.001	
	FEV_1	Measured	137.88			
		Korean model		83.27	<0.001	<0.05
		USA model		72.73	<0.001	

*Choi et al.[4]. [†]Hankinson et al.[2].
SST: total sum of square; SSE: sum of square due to error; FVC: force vital capacity; FEV_1: forced expiratory volume in one second; NS: not significant.

ing Korean models. When compared to the Korean regression models, all SSEs of the USA regression models, but that for FEV_1 in males, were significantly smaller. These facts suggest that these USA regression models are more appropriate as the references for spirometry tests of Koreans. On the contrary, if values estimated with the Korean regression model are used as the reference values, more ventilatory impairment cases would be diagnosed. This is consistent with the study results that the prevalence of ventilatory impairment changed greatly according to the referent model taken[5].

As this study was conducted in a limited region with a relatively small sample size, it is not recommended to generalize the results of this present study to Korean population. Therefore, we did not present any regression model for the spirometric data of this study.

In conclusion, the USA Caucasian regression models showed a better fit to spirometry test data of this study than the Korean regression models did. New regression models with better goodness-of-fits for Korean general population should be introduced.

References

1. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. Eur Respir J 2005;26:319-38.
2. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med 1999;159:179-87.
3. Kim WY, Kim KH, Youn BH, Lee SU, Cho CH, Choi JS, et al. Prediction equations of pulmonary function parameters derived from the forced expiratory spirogram for healthy adults over 50 years old in rural area. Tuberc Respir Dis 1998;45:536-45.
4. Choi JK, Paek D, Lee JO. Normal predictive values of spirometry in Korean population. Tuberc Respir Dis 2005;58:230-42.
5. Oh YM, Hong SB, Shim TS, Lim CM, Koh Y, Kim WS, et al. Effect of a new spirometric reference equation on the interpretation of spirometric patterns and disease severity. Tuberc Respir Dis 2006;60:215-20.