HIGHLIGHTED TOPIC | Pulmonary Physiology and Pathophysiology in Obesity

Physiologically interpretable prediction equations for spirometric indexes

Wojciech Lubiński¹† and Tomasz Gólczewski²
¹Military Institute of Medicine, Central Clinical Hospital of the Ministry of National Defence, Warsaw; and ²Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland

Submitted 26 October 2009; accepted in final form 18 January 2010

Physiologically interpretable prediction equations for spirometric indexes. J Appl Physiol 108: 1440–1446, 2010. First published January 21, 2010; doi:10.1152/japplphysiol.01211.2009.—The need for ethnic-specific reference values of lung function variables (LFs) is acknowledged. Their estimation requires expensive and laborious examinations, and therefore additional use of results in physiology and epidemiology would be profitable. To this end, we proposed a form of prediction equations with physiologically interpretable coefficients: a baseline, the onset age (A0) and rate (S) of LF decline, and a height coefficient. The form was tested with data from healthy, nonsmoking Poles aged 18–85 yr (1,120 men, 1,625 women) who performed spirometry maneuvers according to American Thoracic Society criteria. The values of all the coefficients (also A0) for several LFs were determined with regression of LF on patient’s age and deviation of patient’s height from the mean height in the year group of this patient. S values for forced vital capacity (FVC), forced expiratory volume in 1 s (FEV1), peak expiratory flow, and maximal expiratory flow at 75% of FVC (MEF75) were very similar in both sexes (1.03 ± 0.07%/yr). FEV1/FVC declines four to five times slower. S for MEF25 appeared age dependent. A0 was smallest (28–32 yr) for MEF25 and FEV1. About 50% of each age subgroup (18–40, 41–60, 61–85 yr) exhibited LFs below the mean, and 4–6% were below the 5th percentile lower limits of normal, and thus the form of equations proposed in the paper appeared appropriate for spirometry. Additionally, if this form is accepted, epidemiological and physiological comparison of different LFs and populations will be possible by means of direct comparison of the equation coefficients.

FEV1; height; obesity; lung function decline; reference values

INTERPRETATION OF PULMONARY function testes (PFTs) is usually based on comparisons of data measured in an individual patient with reference predicted values of healthy subjects. As equipment and techniques for lung function testing improve, advanced mathematical models to describe lung function data are implemented. Our intention was to develop a new model for predicted values for spirometry and other measurements, including total lung capacity (TLC), residual volume-to-TLC ratio, and diffusing capacity of the lung for carbon monoxide in the future. Respiratory disorders are important causes of morbidity and mortality all over the world—obstructive pulmonary diseases (as asthma and chronic obstructive pulmonary disease) are expected to be the fourth leading cause of death in 2010 (5) and the third one in 2020 (15). Spirometry is used in diagnosing pulmonary diseases, in the determination of the pulmonary function, and in the response to the therapy. For that reason, interpretation of spirometry should be improved for early detection of obstructive pulmonary diseases. Diagnosis with forced spirometry requires comparison of observed results obtained during the examination with reference values predicted for individuals of the same age, height, and sex (as well as race or ethnicity). As previously suggested (4), it would be best if each laboratory performing lung function tests had its own reference values; however, it is unrealistic. Nevertheless ethnic-specific reference equations should be used whenever possible (17). For these reasons, such prediction equations have been published recently (e.g., Refs. 2, 5, 8, 13, 14).

Elaboration of these equations requires examination of a large population to avoid statistical artifacts. Such extensive examinations necessary to estimate reference values accurately are expensive, laborious, etc. Therefore, it would be profitable if results of the examinations could also be used in other applications different than calculation of predicted values of spirometric indexes for an individual. In particular, it would be profitable if prediction equations proposed by different authors for various populations could be directly comparable in physiological and epidemiological analyses. To make possible comparisons of different populations, sexes, and lung function variables (LFs) by means of a direct comparison of equation coefficients, we propose a physiologically interpretable form of the prediction equations.

The four following criteria were taken into account when such an interpretable form for the regression equations was derived.

1) The form should be interpretable, i.e., equation coefficients should have physiological meaning.
2) There is an onset age A0 for the decline of an LF, i.e., LF declines significantly but not before an average individual is A0 years old.
3) A0 should be determined by means of regression methods in the same way as and together with the other equation coefficients.
4) The form should be as simple as possible to make interpretation easy [e.g., a piecewise linear equation should be preferred if the fraction of the explained variance (R²) for such a form was approximately equal to R² for more sophisticated nonlinear equations].

The third criterion is of special meaning and should be taken into account in the future by all authors even if the form proposed in this paper would not be commonly accepted. The possibility of automatic determinations of values of all the coefficients (also the onset age) with regression was an important criterion for the choice of the equations employed. Currently, other authors who used piecewise linear or nonlinear equations either assumed the same change points for all LFs...
arbitrarily (e.g., Refs. 6, 18) or located those points by means of a combination of graphic analysis to determine the approximate location and $R^2$ to refine the graphic estimation (8). In our opinion, such an arbitrary or manual determination of one coefficient may make automatic determination (with regression) of the other coefficients imprecise.

The second criterion may enable us to express the nonlinear dependence of LFs on age using simple piecewise linear equations (the 4th criterion); however, piecewise nonlinear equations, i.e., equations “linear” in the form but with age-dependent coefficients, should be taken into account if they are significantly better.

It seems to be commonly acceptable that the LF values of an individual are determined by three main factors: the maximally attained values of the variables during early adulthood and the onset age and rate of decline of these values (12). Therefore, the equations of the form proposed in this paper have the following coefficients: 1) a baseline, 2) the onset age of decline, 3) the relative (%/yr) rate of this decline, and 4) a height coefficient describing interpersonal differences not related to age. Despite the fact that weight may influence LF, according to the present tendency, interpersonal differences other than differences in height are ignored (certainly, such categorical variables as sex or ethnicity are taken into account). Although obesity can change some LF values significantly (3, 21), it seems to be treated rather as a “nonpulmonary” pathology of lung function [e.g., compared with asthma in some studies (16)] than an interpersonal difference.

The decline rate and height coefficient may be age dependent. Certainly, the baseline and onset age are constant numbers for a considered LF and sex. Note that prediction equations are elaborated to be used for diagnosing now, and thus not longitudinal but cross-sectional study should be taken into account. Therefore, the above onset age and rate of LF decline are related to both physiological decline according to aging and a birth cohort effect.

The proposed form of the prediction equations was tested using a sample of the Polish population. The following LFs were studied: forced vital capacity (FVC); forced expiratory volume using a sample of the Polish population. The following LFs were studied: forced vital capacity (FVC); forced expiratory volume at 75, 50, and 25% of FVC (MEF75, MEF50, and MEF25, FVC); peak expiratory flow (PEF); maximal expiratory flow at any rate of this LF using simple piecewise linear equations (the 4th criterion); however, piecewise nonlinear equations, i.e., equations “linear” in the form but with age-dependent coefficients, should be taken into account if they are significantly better.

The proposed form of the prediction equations was tested using a sample of the Polish population. The following LFs were studied: forced vital capacity (FVC); forced expiratory volume in one second (FEV1); ratio of FEV1 and FVC (FEV1/FVC); peak expiratory flow (PEF); maximal expiratory flow at 75, 50, and 25% of FVC (MEF75, MEF50, and MEF25, respectively); and forced mid-expiratory flow (MEF75–25).

**MATERIALS AND METHODS**

**Material.** With the permission of the Local Ethics Committee, the proposed form of equations was tested with the use of an already existing database. The database was created in the Military Institute of Health Services (Warsaw, Poland) during performance of the project “Hope for Lungs,” which involved screening of spirometric examinations in Polish volunteers without any exclusion criteria to ensure the diagnosis of early obstructive lung disease. It was conducted from 2002 to 2005 by the Military Institute of Health Services. With a mobile laboratory, the examinations were performed at 93 sites, including both large cities and villages throughout Poland. For all subjects, after an interview and written consent were obtained, an examination was performed in a sitting position using a regularly calibrated spirometer (LungTEST1000 by MES, Poland). Tests took place during the summer months (between June and September) from 9 AM to 4 PM.

In total, 5,130 women and 4,716 men were examined. American Thoracic Society guidelines (1) were used for the technical evaluation of the measurements. All examinations were performed and analyzed by the same group of six qualified employees (the author and other physicians and nurses performing routine spirometry in the Central Clinical Hospital of the Ministry of National Defence). The final selection was performed by both authors of this paper, taking into account medical and technical criteria. The results of the examinations were excluded for several subpopulations (exclusion criteria): 1) persons younger than 18 yr or older than 85 yr of age (too small number of participants), 2) smokers, subjects diagnosed with chronic obstructive pulmonary disease and asthma, or individuals reporting the occurrence of chronic cough or dyspnea within the last 12 months (the standard criteria when prediction equations are elaborated), and 3) individuals who were unable to undergo correct spirometry (the database used contains such results for other applications). Due to these criteria, 3,505 women and 3,596 men were rejected. Consequently, the results for 1,625 women and 1,120 men were used in the analysis. Table 1 presents details concerning the groups of individuals. Note that the elderly group was relatively large.

**Study design.** Using the database from the Military Institute of Health Services, Warsaw, Poland, the following were determined.

1) (For the whole range of age) the mean height for each year subgroup was determined by regression of height on age.

2) (For the whole range of age) physiologically interpretable coefficients of prediction equations were determined by regression of LFs on age and the deviation of subject’s height from this mean height.

3) (For three age groups separately) the accuracy of the established equations was tested by means of analysis of a) statistical significance of the mean difference between predicted and observed values of LFs, b) percentages of subjects having LFs below their predicted value or below the lower limit of normal (LLN) for their sex, age, and height (the online supplementary material contains the same analysis for equations proposed by other authors).

**Data analysis.** The following form of prediction equations was accepted (details of the form derivation are presented in the online supplementary materials):

$$LF_m = m_{LF_m} \cdot \left\{1 + 0.5[abs(A - A_{0LF}) + (A - A_{0LF})] \cdot S_{LF_m} + h_{LF_m} \cdot \Delta H \right\} (I)$$

where A is subject’s age; $A_{0LF}$ is the onset age of decline of a particular LF in an analyzed population; $S_{LF_m}$ is the rate of this decline; $abs(A - A_{0LF})$ is the absolute value of $(A - A_{0LF})$; $\Delta H$ is the deviation of the height (H) of an examined subject from the mean height for his/her year group $[Ho(A)]$, i.e., $\Delta H = H - Ho(A)$ where the

| Table 1. Reference sample selection criteria and characteristics of the study sample |
|-----------------|-----------------|
| **Participants** | **Men** | **Women** |
| Total | 4,716 | 5,130 |
| Less aged <18 or >85 yrs | 4,293 | 4,767 |
| Less ever smokers | 2,385 | 3,382 |
| Less invalid or irreproducible data | 1,454 | 1,825 |
| Less unhealthy subjects | 1,120 | 1,625 |
| Age in years | | |
| 18–45 | 311 | 316 |
| 46–65 | 472 | 766 |
| 65–85 | 337 | 543 |
| Age, yr | 55 ± 16 | 57 ± 14 |
| Height, cm | 173 ± 7 | 161 ± 6 |
| FEV1, liters | 3.4 ± 1.0 | 2.4 ± 0.7 |
| FVC, liters | 4.5 ± 1.2 | 3.1 ± 0.8 |

Data are presented as the number of participants ($n$) or mean ± SD. FEV1, forced expiratory volume in 1 s; FVC, forced vital capacity.
function \( H_0(A) \) was found by means of regression of height on age; \( h_{LF,m} \) is the height coefficient; \( m_{LF,m} \) is some kind of baseline, i.e., the value of a considered LF for the average 18-yr-old subject of height equal to the mean height for this year group, i.e., of height \( H = H_0 \) (18 yr); \( LF_m \) is the predicted value of the considered LF for an examined subject (the mean value for all subjects of height and age equal to height and age of this subject).

It should be stressed that the form of Eq. 1 enables us to determine the onset age \( A_0_{LF} \) together with the other coefficients by means of regression (see online supplementary material).

**Equation 1** means that if the subject's age is less than \( A_0_{LF} \) then

\[
LF_m = m_{LF,m} \cdot (1 + h_{LF,m} \cdot \Delta H)
\]

otherwise

\[
LF_m = m_{LF,m} \cdot (1 + (A - A_0_{LF}) \cdot S_{LF,m} + h_{LF,m} \cdot \Delta H)
\]

**Equation 1**

Note the following.

1) \( \Delta H \) is the residual from the regression of height on age, in fact. Hence it appears that the height \( H \) is proposed to be decomposed into the following two parts: \( H_0(A) \) connected with age and \( \Delta H \) representing age-independent interpersonal differences in a subject's year group. The use of \( \Delta H \) instead of the height results in the influence of age being completely separated from the influence of interpersonal differences in height. Reasons for replacing height with its deviation from the mean for the subject's year group are detailed in the online supplementary material.

2) \( A_{0,LF} \) and \( S_{LF,m} \) describe differences in a whole population seen as a snapshot (the cross-sectional study) related to age treated as an index characterizing an individual. Thus they reflect all physiological changes caused by ageing (including influence of age through influence on height) and differences in generations (a birth cohort effect including differences in the mean height between generations when their participants were 18 yr old).

3) \( h_{LF,m} \) describes relative rate, and thus 1/yr (or%/yr) is its unit. Analogously, 1/cm is the unit of \( h_{LF,m} \).

The equation for the regression of height on age, required to calculate \( \Delta H \), is analogous (the reason is explained in the online supplementary material):

\[
H_0(A) = m_{LF} \cdot [1 + 0.5 \cdot \text{abs}(A - A_0_{LF}) + (A - A_0_{LF}) \cdot S_{LF,m}] (4)
\]

where \( m_{LF} \) is the mean height of subjects 18 yr old, \( A_0_{LF} \) is the onset age of height decline, \( S_{LF,m} \) is the rate of this decline (the above note for \( A_{0,LF} \) and \( S_{LF,m} \) relates also to \( A_0_{LF} \) and \( S_{LF,m} \)).

The decline rate for MEF25 depends significantly on age, and therefore \( S_{MEF25,m} \) is expressed by the following formula:

\[
S_{MEF25,m}(A) = S_0_{MEF25,m} + S_{1,MEF25,m} \cdot (A - A_{0,MEF25}) (5a)
\]

For the other LFs, \( S_{LF,m} \) could be assumed to be independent of age

\[
S_{LF,m}(A) = S_0_{LF,m} (5b)
\]

because coefficients \( S_{1,LF,m} \) were statistically insignificant or their neglect reduced \( R^2 \) by <1%.

The lower limit of normal (LLN) was determined using a method proposed by Healy et al. (10), which has been used by other authors (2, 6). According to that method, the fifth percentiles of residuals from regression for mean were then determined for consecutive age groups. These percentiles were then regressed against the mean ages of the corresponding groups using the following equation:

\[
LF_{5\%} = m_{LF,5\%} \cdot \{1 + 0.5 \cdot \text{abs}(A - A_{0,LF}) + (A - A_{0,LF}) \cdot S_{LF,5\%}\} (6a)
\]

LLN of a LF is the sum of \( LF_m \) and \( LF_{5\%} \), which can be expressed by the equation having the form as Eq. 1:

\[
LF_{LLN} = m_{LF,LLN} \cdot \{1 + 0.5 \cdot \text{abs}(A - A_{0,LF}) + (A - A_{0,LF}) \} \cdot S_{LF,LLN} + h_{LF,LLN} \cdot \Delta H \} (6b)
\]

where \( m_{LF,LLN}, S_{LF,LLN}, \) and \( h_{LF,LLN} \) have the same meaning as \( m_{LF,m}, S_{LF,m}, \) and \( h_{LF,m} \), however, their values are different (see the online supplementary material for details of LLN determination).

The standard statistical calculations, such as regression and confidence intervals, were performed with the computer system Statistica (StatSoft). Other calculations were performed using our own computer programs.

**RESULTS**

Table 2 presents the values of the equation coefficients describing the dependence of LFs (their predicted values and LLN) on age and height. For example, the predicted value of FEV1/FVC and its LLN for an 80-year-old woman should be calculated as follows [note, that dependence of FEV1/FVC on height appeared to be not statistically significant in women; such a weak dependence was observed by other authors and even ignored a priori (8, 18)]:

\[
\begin{align*}
\text{FEV1/FVC}_{m} &= 81.174 \cdot [1 + (80 - 27.5548) \cdot (-0.00209)] \\
&= 72.3 \\
\text{FEV1/FVC}_{LLN} &= 70.8010 \cdot [1 + (80 - 27.5548) \cdot (-0.00251)] \\
&= 61.5
\end{align*}
\]

These results are almost identical to those showed by Hardie et al. (9) for Norwegian elderly.

In the case of other LFs, the height deviation \( \Delta H \) should be calculated first. For example, for the above-described woman (height 155 cm):

\[
H_0(80) = 164.503 \cdot [1 + (80 - 35.4491) \cdot (-0.001)]
\]

\[
= 157.174,
\]

and thus \( \Delta H \) would be equal to:

\[
\Delta H = H - H_0(80) = 155 - 157.174 = -2.174 \text{ cm}
\]

Then, the predicted value or LLN should be calculated, as in the case of FEV1/FVC; however, the height term equal to \((-2.174 \cdot h_{LF,m})\) or \((-2.174 \cdot h_{LF,LLN})\) should be added according to Eq. 3 or 6b.

The age of onset of LF decline \( A_{0,LF} \) is different for different LFs (Table 2). For both sexes, MEF25 and FEV1 start to decrease first, whereas PEF and MEF75 decrease last.

The decline rate appeared to be age dependent only in the case of MEF25. The rates of FVC, FEV1, PEF, and MEF75 decline in both sexes were very similar (1.03 ± 0.07%/yr for the mean). The rates of decline of middle airflows (i.e., MEF50 and MEF75–25) in both sexes were also similar (1.37 ± 0.07%/yr for the mean). MEF25 initially declined faster (2.0–2.2%/yr for the mean), but that decline decelerated according to the age-dependent component, i.e., to 0.00012\(A\)-32.1) for males and 0.00016\((A-28.5)\) for females (Table 2, Fig. 1). FEV1/FVC decreased much slower than the other LFs.

As data presented in Tables 3–5 demonstrate, equations of the proposed form accurately describe the dependences of LFs on age. In particular, despite the fact that the coefficients' values were found with regression performed for the whole age range (18–85 yr), the equations provided correct results for all three age groups: the young (18–40 yr), middle aged (41–60 yr), and elderly (61–85 yr) treated separately. The mean difference between the observed and predicted values differed insignificantly from zero in all the age subgroups since the
Table 2. Coefficients of the prediction equations for the mean lower limit of normal

<table>
<thead>
<tr>
<th>m [#]</th>
<th>S [1/yr]</th>
<th>A0 [yrs]</th>
<th>h [1/cm]</th>
<th>R²</th>
<th>RSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height [cm]</td>
<td>164.503</td>
<td>35.4491</td>
<td>-0.0010</td>
<td>0.12</td>
<td>5.82</td>
</tr>
<tr>
<td>FEV1/FVC [%]</td>
<td>81.1741, 70.8010</td>
<td>27.5548</td>
<td>-0.00209, -0.00251</td>
<td>ns</td>
<td>0.14</td>
</tr>
<tr>
<td>FVC [L]</td>
<td>4.12789, 3.10154</td>
<td>31.4104</td>
<td>-0.00992, -0.01075</td>
<td>0.00843, 0.01122</td>
<td>0.60</td>
</tr>
<tr>
<td>FEV1 [L]</td>
<td>3.36521, 2.51694</td>
<td>28.9970</td>
<td>-0.01070, -0.01209</td>
<td>0.00741, 0.00991</td>
<td>0.61</td>
</tr>
<tr>
<td>MEF75 [L/s]</td>
<td>9.03507, 5.82829</td>
<td>39.2272</td>
<td>-0.01293, -0.01454</td>
<td>ns</td>
<td>0.38</td>
</tr>
<tr>
<td>FEV1/FVC [%]</td>
<td>7.33389, 5.01139</td>
<td>35.7733</td>
<td>-0.00973, -0.01231</td>
<td>0.00740, 0.01083</td>
<td>0.35</td>
</tr>
<tr>
<td>PEF [L/s]</td>
<td>6.50329, 4.26791</td>
<td>35.4849</td>
<td>-0.01044, -0.01535</td>
<td>0.00711, 0.01083</td>
<td>0.32</td>
</tr>
<tr>
<td>MEF25 [L/s]</td>
<td>4.93453, 2.75249</td>
<td>33.0000</td>
<td>-0.01349, -0.01687</td>
<td>0.00641, 0.01192</td>
<td>0.34</td>
</tr>
<tr>
<td>MEF50 [L/s]</td>
<td>3.61429, 1.97036</td>
<td>30.0000</td>
<td>-0.02185, -0.02885</td>
<td>0.00409, 0.00707</td>
<td>0.48</td>
</tr>
<tr>
<td>MEF25 [L/s]</td>
<td>3.61429, 1.97036</td>
<td>30.0000</td>
<td>-0.02185, -0.02885</td>
<td>0.00409, 0.00707</td>
<td>0.48</td>
</tr>
</tbody>
</table>

LLN, lower limit of normal; m, the mean/LLN for 18-yr-old subjects of the mean height (female 164.5 cm, male 179.8 cm); # — the unit as for the corresponding lung function variable; A0, the onset age of decline; S, the rate of the decline with age; h, height coefficient; R², fraction of explained variance (regression for the mean); RSD, residual SD, *coefficient for the quadratic component (P < 0.01 for this component, P < 0.001 for the other showed coefficients; ns, coefficients statistically insignificant, i.e., P > 0.05).

DISCUSSION

The subset of the database we used to test the proposed form of spirometric prediction equations was comprised of healthy Polish, nonsmoking volunteers. The exclusion criteria were in accordance with the published ATS recommendations (1), and their accuracy in population selection was confirmed in 2007 by Johannessen (11). Although the number of subjects was not significantly different from those described by other authors (2, 6, 8, 13, 14), the age distribution was different. Our sample was approximately a cross-section of the Polish population with respect to all factors except age: the elderly subgroup was larger than it seemingly should be. However, this is not an imperfection in determining prediction equations. Indeed, the mathematical methods of regression determine the coefficients of an equation by minimizing the sum of squared differences between predicted and observed values. Therefore, the coefficients are best fitted to the largest age subgroup, as it has the greatest influence on this sum. To distinguish pathological lesions from involution, the predicted values for the elderly should be determined with particular care. Hence, it appears that the age distribution should be at least uniform. For that reason, the age distribution of our sample was approximately uniform (Table 1). Additionally, the proposed form of equations requires an adequately high number of I) the young to determine accurately the baseline mLF.m, 2) middle aged to
determine the onset age $A_{0LF}$, and 3) the elderly to determine accurately the decline rate $S_L F_{m}$. The accuracy of the proposed form was proven by the adequate percentages of the population with LFs lower than the mean and LLN in all of the three age groups analyzed separately (Tables 3–5). The percentage of the population with LFs lower than the 50th percentile (i.e., the median) was even closer to 50%; however, only the mean is presented herein because the 50th percentile is not commonly accepted to be the predicted value (despite the practice of using the 5th percentile as LLN). When the regression with an equation of a firm form is performed for the whole range of age, it has to give correct results if the whole range is considered (e.g., the mean difference between all predicted and observed values should be statistically insignificant). However, if the firm form of the regression equation is not appropriate, the results may be incorrect if subranges are considered (e.g., the mean difference between predicted and observed values in the young may be significantly smaller than zero, whereas in the elderly, significantly greater than zero despite that the difference for the whole range is not significantly different from zero because the underestimation in the young compensates the overestimation in the elderly). Since our equations appeared appropriate in subranges of age (Tables 3–5), our physiologically form of equations seems to be proper. Note additionally that the fraction of explained variance ($R^2$) (Table 2) was very close to those obtained by other authors (2, 6, 8, 13, 14), e.g., $R^2$ values obtained by Falaschetti et al. (6) were slightly greater for women and smaller for men (the online supplementary material contains brief comparison between our equations and those proposed by some other authors related to the studied sample of the Polish population).

In contrast to other authors [with the exception of Quanjer et al. (18)], we proposed to express the nonlinearity of LF dependence on age using piecewise linear equations (only MEF25 required a quadratic form; Fig. 1). However, contrary to Quanjer et al., we determined separately the change point from measurements for each individual LF. On the other hand and in contrast to Hankinson et al. (8), we determined change points automatically, i.e., together with the other coefficients by means of regression methods, while those authors located the points by means of a combination of graphic analysis to determine the approximate location and $R^2$ to refine the graphic estimation. Perhaps our mathematical method of point location was the main reason that we could express the nonlinearity with sufficient accuracy using piecewise linear equations, while other authors had to use nonlinear or even piecewise nonlinear (e.g., Ref. 6) equations.

Although it is still suggested that at least the FEV1 decline accelerates with age (19), such acceleration was not observed herein (Fig. 2). Indeed, if such acceleration existed and the mean was determined adequately with a linear equation, the percentage of subjects with FEV1 lower than the inadequate mean value would be greater in the elderly than in the middle aged. In this study, the FEV1 for 48% of all middle aged, 48% of elderly women, and 50% of elderly men was below the mean described with linear equations (Tables 3 and 4). Hence, it appears that FEV1 declines at an approximately constant rate. The above is also true for the other LFs (except MEF25, as its decline accelerates with age). Note that our results of the cross-sectional study (decline with age) are not contradictory to results related to the decline rate analyzed in longitudinal studies (decline with ageing) suggesting acceleration (e.g., Ref.

### Table 3. Mean differences between observed values and those predicted with prediction equations for the three age groups of men

<table>
<thead>
<tr>
<th>Age Range</th>
<th>18–40 yr</th>
<th>41–60 yr</th>
<th>61–85 yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV1/FVC [%]</td>
<td>−0.170 [−0.88; 0.54]</td>
<td>0.066 [−0.49; 0.62]</td>
<td>−0.028 [−0.61; 0.56]</td>
</tr>
<tr>
<td>FVC [L]</td>
<td>−0.039 [−0.14; 0.06]</td>
<td>0.011 [−0.05; 0.08]</td>
<td>−0.002 [−0.06; 0.06]</td>
</tr>
<tr>
<td>FEV1 [L]</td>
<td>−0.030 [−0.12; 0.05]</td>
<td>0.007 [−0.05; 0.07]</td>
<td>−0.003 [−0.06; 0.05]</td>
</tr>
<tr>
<td>MEF75–25 [L/s]</td>
<td>0.011 [−0.17; 0.19]</td>
<td>−0.006 [−0.12; 0.11]</td>
<td>−0.011 [−0.09; 0.07]</td>
</tr>
<tr>
<td>PEF [L/s]</td>
<td>0.006 [−0.19; 0.19]</td>
<td>−0.029 [−0.18; 0.12]</td>
<td>0.027 [−0.14; 0.20]</td>
</tr>
<tr>
<td>MEF75 [L/s]</td>
<td>0.001 [−0.22; 0.22]</td>
<td>−0.014 [−0.19; 0.16]</td>
<td>0.012 [−0.16; 0.19]</td>
</tr>
<tr>
<td>MEF50 [L/s]</td>
<td>−0.011 [−0.22; 0.21]</td>
<td>0.007 [−0.14; 0.16]</td>
<td>−0.009 [−0.12; 0.11]</td>
</tr>
<tr>
<td>MEF25 [L/s]</td>
<td>−0.019 [−0.11; 0.07]</td>
<td>0.004 [−0.04; 0.05]</td>
<td>0.017 [−0.01; 0.05]</td>
</tr>
</tbody>
</table>

Diff, the mean difference; CI, the 95% confidence interval; %, the percentage of the age group with the value of a lung function variable below the mean value estimated with the proposed equation. MEF75–25, forced mid expiratory flow; PEF, peak expiratory flow and maximal expiratory flow at 75, 50, and 25% of FVC (MEF75, MEF50, and MEF25, respectively).

### Table 4. Mean differences between observed values and those predicted with prediction equations for the 3 age groups for women

<table>
<thead>
<tr>
<th>Age Range</th>
<th>18–40 yr</th>
<th>41–60 yr</th>
<th>61–85 yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV1/FVC [%]</td>
<td>−0.006 [−0.81; 0.80]</td>
<td>−0.040 [−0.47; 0.39]</td>
<td>0.048 [−0.39; 0.49]</td>
</tr>
<tr>
<td>FVC [L]</td>
<td>−0.026 [−0.10; 0.04]</td>
<td>0.012 [−0.03; 0.05]</td>
<td>−0.004 [−0.04; 0.03]</td>
</tr>
<tr>
<td>FEV1 [L]</td>
<td>−0.013 [−0.07; 0.05]</td>
<td>0.004 [−0.03; 0.04]</td>
<td>−0.001 [−0.03; 0.03]</td>
</tr>
<tr>
<td>MEF75–25 [L/s]</td>
<td>−0.031 [−0.19; 0.12]</td>
<td>0.008 [−0.05; 0.07]</td>
<td>−0.002 [−0.04; 0.04]</td>
</tr>
<tr>
<td>PEF [L/s]</td>
<td>−0.123 [−0.29; 0.05]</td>
<td>0.028 [−0.08; 0.14]</td>
<td>0.006 [−0.08; 0.10]</td>
</tr>
<tr>
<td>MEF75 [L/s]</td>
<td>−0.135 [−0.30; 0.03]</td>
<td>0.055 [−0.05; 0.16]</td>
<td>−0.014 [−0.11; 0.08]</td>
</tr>
<tr>
<td>MEF50 [L/s]</td>
<td>−0.043 [−0.18; 0.10]</td>
<td>0.014 [−0.07; 0.01]</td>
<td>−0.001 [−0.07; 0.06]</td>
</tr>
<tr>
<td>MEF25 [L/s]</td>
<td>−0.003 [−0.08; 0.07]</td>
<td>0.005 [−0.02; 0.03]</td>
<td>0.023 [0.00; 0.04]</td>
</tr>
</tbody>
</table>
(12) because our $S_{\text{LF}_m}$ describe together both ageing and the birth cohort effect.

The proposed form of the equations enabled us to compare the coefficients for all LFs, even between both sexes. The $m_{\text{LF}}$ coefficient is a kind of the constant term in the equation. Such terms in the equations of other authors have no physiological meaning. Here, $m_{\text{LF}}$ is the value of the corresponding LF in average young persons of height equal to the mean height for this year group, and therefore $m_{\text{FEV1}}/m_{\text{FVC}}$ and $m_{\text{MEF25}}/m_{\text{MEF75}}$ for both the mean and LLN. As expected, for each measurable LF (i.e., except the calculated FEV1/FVC), the $m_{\text{LF}}$ for women is smaller than that for men. Of interest:

$$\frac{m_{\text{FEV1.LLN}}}{m_{\text{FVC.LLN}}} = m_{\text{FEV1/FVC}}$$

Thus it appears that healthy young subjects have the FEV1/FVC value equal to the mean, on average, even if their FEV1 and FVC are at LLN levels.

The fact that $A_0$ for all LFs is greater than 18 yr confirms the suggestions accepted by many authors (beginning from Ref. 18) that LFs weakly change with age in young adults (12, 20). Note, however, that the FEV1 and FVC values in men have a maximum for age equal to 30 yr if subjects of a constant height (175 cm in Fig. 2) are considered in a cross-sectional study.

Different values of $A_{0\text{LF}}$ obtained for different LFs seem to explain why various authors suggest different values for the onset age of the decline: from 25 yr (18) to 40 yr (7, 20). The age of onset for FEV1/FVC decline depends on when FEV1 and FVC start to decrease. If FVC starts to decrease later than FEV1, then FEV1/FVC begins to decline simultaneously with FEV1 (as in women). If FEV1 and FVC start to change simultaneously (as in men), then FEV1/FVC begins to change later (since FEV1 decreases faster than FVC by $-0.1\%$ per year, FEV1/FVC must decline sooner or later). The reasons explaining the above differences between sexes are not clear. Future investigations, especially those performed by other authors for other populations, should show whether this difference is a statistical artifact or an actual fact. Note, however, that the absolute differences between sexes in $A_{0\text{FEV1/FVC}}$ and $A_{0\text{H}}$ are almost the same, which might suggest some connection. Since such a small value of $A_{0\text{H}}$ in men seems to be mainly caused by the birth cohort effect, $A_{0\text{FEV1/FVC}}$ and $A_{0\text{H}}$ values may be different in other ethnic groups of another political and economic history.

The role of the forced expiratory flow-volume curve in diagnostics is founded on the principle that its general shape is similar for all healthy subjects, regardless of age. However, there are two differences between the curves for the young and elderly: 1) the curves for the elderly are smaller than those for the young; 2) the curves are more concave for the elderly. This means that changes in PEF, MEF75, FVC, and FEV1 should be proportional (shape similarity); however, MEF50 and MEF25 should decrease slightly faster and earlier (concavity). The values obtained for decline rates ($S_{\text{LF}_m}$) correspond exactly to the above conditions (Table 2).

Indeed, as $S_{\text{LF}_m}$ is the relative rate, the equality of its values for PEF, MEF75, FVC, and FEV1 ($1.03 \pm 0.07%/yr$) indicates proportional decreases of these LFs. A slightly greater $S_{\text{LF}_m}$ for

<table>
<thead>
<tr>
<th>Percentage of age groups with a lung function variable below the lower limit of normal estimated with the proposed equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>FEV1/FVC</td>
</tr>
<tr>
<td>FVC</td>
</tr>
<tr>
<td>FEV1</td>
</tr>
<tr>
<td>MEF75–25</td>
</tr>
<tr>
<td>PEF</td>
</tr>
<tr>
<td>MEF75</td>
</tr>
<tr>
<td>MEF50</td>
</tr>
<tr>
<td>MEF25</td>
</tr>
</tbody>
</table>

Fig. 2. Predicted values of forced vital capacity (FVC) and forced expiratory volume in 1 s (FEV1) in men (height 175 cm) and women (height 160 cm) by age. Solid lines, prediction equations from this study; dotted lines, from Falaschetti et al. (6); black lines, the mean; gray lines, lower limit of normal; thin lines, predictions beyond the age range for which equations were elaborated; light gray dots, actual results for the examined individuals.
MEF50 and MEF75–25 and an initial two-times greater $S_{M_{EP25m}}$ cause the curve to become concave.

Prediction equations proposed by different authors for various populations are different. In particular, those equations seem to be inappropriate for Polish population (especially for the Polish elderly, see tables in the online supplementary material). Our equations describe the studied sample much better than other equations; however, other authors have fitted coefficients of their own equations to their own populations. Therefore, to compare directly only the forms of equations, we performed regressions for our sample using the exponential-quadratic form of Falaschetti et al. (6), Brandli et al. (2), and Langhammer et al. (14), the additive quadratic form of Hankinson et al. (8), and the following multiplicative quadratic form: $LF = (A_0 + A_1 \cdot age + A_2 \cdot age^2) \cdot height^{A_3}$. The values of $R^2$ for the exponential-quadratic form were smaller by 0.01 to 0.06 than those shown in Table 2, whereas the $R^2$ values for the two next forms were exactly the same as in Table 2 for all LFs (considering two decimal places). If prediction equations of the form used by Hankinson et al. fit to our population, results are usually (especially in men) as good as those obtained with our form (see tables in the online supplementary material). Hence it appears that our form of equations can be both physiologically interpretable and appropriate for prediction equations as other forms or even better than some of them.

If all prediction equations had the interpretable form proposed in this paper, it would be possible to state whether the decline rate or the onset age or other factors differ between two populations. For example, differences in the function $H_0(A)$ and similar values of $h_{LF,m}$ would suggest general similarity of LF dependence on height and differences between populations related to different economic, historical, etc. conditions. On the other hand, differences in $h_{LF,m}$ might suggest differences in body proportions between populations (e.g., American-Africans on average have a smaller trunk-to-leg ratio than do Caucasians, and thus differences in FEV1 and FVC between those races (8) could probably be explained by means of direct analysis of physiologically interpretable parameters).

Such analysis is impossible when another form is used. For example, no conclusion can be drawn from comparison of the coefficients of prediction equations having the additive quadratic form for Polish and American Caucasian populations (the online supplementary material).

The above physiological considerations are only an illustration that epidemiological studies necessary to elaborate prediction equations for spirometric indexes will only connect with physiological findings if our form of the equations is accepted. Other parameters, such as lung volumes or diffusing capacity of the lung for carbon monoxide, which are now an integral part of a comprehensive functional evaluation of patients suffering from cardiopulmonary diseases, could not be examined because of no corresponding data in the database. Nevertheless, we suggest that they should be investigated in future studies with similar analysis.

In conclusion, a physiologically interpretable form of prediction equations may be interesting to examine since, on the one hand, it is at least as appropriate as the other forms. On the other hand, it enables us to combine physiological knowledge with diagnostic practice, as well as it enables epidemiologists to compare various populations by direct analysis of equation coefficients. If change points are determined by regression together with the other coefficients, the nonlinear age dependence of almost all lung function variables can be accurately described by piecewise linear equations.

ACKNOWLEDGMENTS

The authors thank B. Stankiewicz from the Institute of Biocybernetics and Biomedical Engineering, PAS, for assistance with the initial data analysis, as well as M. D. Tadeusz Plusa from the Military Institute of Health Services, Warsaw, Poland, for scientific support.

DISCLOSURES

No conflicts of interest are declared by the authors.

REFERENCES