MULTI-ETHNIC REFERENCE VALUES FOR SPIROMETRY FOR THE 3-95 YEAR AGE RANGE: THE GLOBAL LUNG FUNCTION 2012 EQUATIONS

Report of the Global Lung Function Initiative, ERS Task Force to establish improved Lung Function Reference Values, endorsed by the ATS, ANZSRS, TSANZ, APSR and the ACCP

Endorsed by the European Respiratory Society (ERS), American Thoracic Society (ATS), Australian and New Zealand Society of Respiratory Science (ANZSRS), Asian Pacific Society for Respirology (APSR), Thoracic Society of Australia and New Zealand (TSANZ) and the American College of Chest Physicians (ACCP).

Philip H. Quanjer¹, Sanja Stanojevic², Tim J. Cole³, Xaver Baur⁴, Graham L. Hall⁵, Bruce Culver⁶, Paul L. Enright⁷, John L. Hankinson⁸, Mary S.M. Ip⁹, Jinping Zheng¹⁰, Janet Stocks¹¹ and the ERS Global Lung Function Initiative (GLI)

The ERS Global Lungs Initiative (see www.lungfunction.org):

Chairs:
J. Stocks, X. Baur, G.L. Hall, B. Culver

Analytical team:
P.H. Quanjer, S. Stanojevic, T.J. Cole, J. Stocks

Additional members of steering committee: J.L. Hankinson, P.L. Enright, J.P. Zheng, M.S.M. Ip

Statistical reviewer:
C. Schindler

Persons and centres contributing data to this manuscript:
O.A. Al-Rawas, Department of Medicine, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Sultanate of Oman; H.G.M. Arets; Department of Pediatric Pulmonology, Wilhelmina Children’s Hospital, University Medical Center Utrecht, Utrecht, The Netherlands; C. Bărbăra, The Portuguese Society of Pneumology, Lisbon, Portugal; R.G. Barr, Columbia University Medical Center, New York, NY, USA and the MESA study; E. Bateman, University of Cape Town Lung Institute, Cape Town, South Africa; C.S. Beardsmore, Department of Infection, Immunity and Inflammation (Child Health), University of Leicester, Leicester, UK; H. Ben Saad, Laboratory of Physiology, Faculty of Medicine, Sousse, University of Sousse, Tunisia; B. Bruneckreif, Institute for Risk Assessment Sciences, Universiteit Utrecht, Utrecht, the Netherlands; P.G.J. Burney, National Heart and Lung Institute, Imperial College, London; R.B. Dantes, Philippine College of Chest Physicians, Manila, Philippines; W. Dejsomritrat, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand; D. Dockery, Department of Environmental Health, Department of Epidemiology, Boston, MA, USA; H. Eigen, Section of Pulmonology and Intensive Care, James Whitcomb Riley Hospital for Children, Indiana University School of Medicine, Indianapolis, IN, USA; A. Falaschetti, [Health Survey for England 1995-1996 (HSE)], International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College, UK; B. Fallon, Respiratory Laboratory, Nepean Hospital, Penrith, Australia; A. Fulambarker, Pulmonary Division, Rosalind Franklin University of Medicine and Science, The Chicago Medical School, Chicago, IL, USA; M. Gappa [LUNOKID study group], Children’s Hospital and Research Institute, Marienhospital Wesel, Germany; M.W. Gerbase, Division of Pulmonary Medicine, University Hospitals of Geneva, Geneva, Switzerland, and the SAPALDIA cohort study; T. Gislason, Landsþipstali University Hospital, Dept. of Allergy, Respiratory Medicine and Sleep, Reykjavik, Iceland; M. Golshan, Bamdad Respiratory Research Institute, Isfahan, Iran; C.J. Gore, Physiology Department, Australian Institute of Sport, Belconnen, Australia; A. Gulsvik, Department of Thoracic Medicine, Institute of Medicine, University of Bergen, Bergen, Norway; G.L. Hall, Respiratory Medicine, Princess Margaret Hospital for Children, Perth, Australia; J.L. Hankinson, [NANES, NANES III Special data sets], Hankinson Consulting, Valdosta, GA, USA; A.J. Henderson, [ALSPAC, http://www.bris.ac.uk/alspac], University of Bristol, Bristol, UK; E. Hnizdo, Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, Morgantown, WV, USA [NANES IV]; M.S.M. Ip, Dept. of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China; C. Janson, Department of Medical Sciences: Respiratory Medicine & Allergology, Uppsala University, Sweden; C. Jenkins, Woolcock Institute of Medical Research, Sydney, Australia; A. Jithoo, University of Cape Town Lung Institute, Cape Town, South Africa; S. Karrasch, Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Hospital of the Ludwig-Maximilians-University, Munich, Germany [KORA study]; G.S. Kerby (Lung Function Measures in Preschool Children with Cystic Fibrosis study

Corresponding author:
Philip Quanjer, e-mail: pquanjer@gmail.com
1 NUMBERS OF SUBJECTS IN ORIGINAL AND FINAL GROUPS

Of the datasets from 33 countries that were shared with the Global Lung Function Initiative, 26 could be included in the four groups that were eventually formed. The original number of subjects, i.e., before exclusion due to missing data or outliers, are presented in Table E1. Of the Mexican data, those in children and adolescents from Mexico City could not be included because the high predicted values did not fit in any of the groups.

Table E1 – Overview of countries sharing data with the Global Lung Function Initiative. The numbers below include data that were subsequently excluded for reasons outlined in the printed text and in Table E2.

<table>
<thead>
<tr>
<th>Country</th>
<th>N</th>
<th>Country</th>
<th>N</th>
<th>Country</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>273</td>
<td>Netherlands</td>
<td>3319</td>
<td>China</td>
<td>3483</td>
</tr>
<tr>
<td>Austria</td>
<td>982</td>
<td>Norway</td>
<td>1535</td>
<td>France</td>
<td>63376</td>
</tr>
<tr>
<td>Austria</td>
<td>333</td>
<td>Poland</td>
<td>220</td>
<td>India</td>
<td>2548</td>
</tr>
<tr>
<td>Brazil</td>
<td>178</td>
<td>Portugal</td>
<td>137</td>
<td>Iran</td>
<td>6137</td>
</tr>
<tr>
<td>Canada</td>
<td>329</td>
<td>Sweden</td>
<td>123</td>
<td>Oman</td>
<td>1256</td>
</tr>
<tr>
<td>Chile</td>
<td>102</td>
<td>Switzerland</td>
<td>11756</td>
<td>Pakistan</td>
<td>2928</td>
</tr>
<tr>
<td>China</td>
<td>5114</td>
<td>Taiwan</td>
<td>2806</td>
<td>Philippines</td>
<td>316</td>
</tr>
<tr>
<td>Germany</td>
<td>4708</td>
<td>Thailand</td>
<td>3262</td>
<td>South Africa</td>
<td>146</td>
</tr>
<tr>
<td>Iceland</td>
<td>164</td>
<td>Tunisia</td>
<td>870</td>
<td>Total</td>
<td>80190</td>
</tr>
<tr>
<td>Israel</td>
<td>124</td>
<td>UK</td>
<td>16888</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>1818</td>
<td>Uruguay</td>
<td>156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>2252</td>
<td>USA</td>
<td>18212</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>4236</td>
<td>Venezuela</td>
<td>243</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total subjects used in final analyses: 80140

Dataset not used in final analyses

<table>
<thead>
<tr>
<th>Country</th>
<th>N</th>
<th>Country</th>
<th>N</th>
<th>Country</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total: 80140

After leaving out various datasets from the regression analysis (Table E2) for reasons delineated in the printed manuscript, the breakdown of numbers by age in each of the four groups is as in Table E3.

2 MODELLING SPIROMETRIC INDICES

The LMS method, implemented in the GAMLSS package [1] in the statistical software R [Version 2.14.1; R Foundation, http://www.r-project.org] allows modelling the expected mean (μ or M), the coefficient of variation (σ or S), and skewness (λ or L). A continuous, smooth fit over the entire age range is obtained by the use of splines. Applying the methodology described by Cole et al. [2] the best fit was estimated using untransformed

Table E2 – Number of subjects that could not be included in the final analyses for reasons delineated in the printed manuscript.

<table>
<thead>
<tr>
<th>Data submitted</th>
<th>160,330</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown ethnicity</td>
<td>63,865</td>
</tr>
<tr>
<td>Suspected asthma</td>
<td>805</td>
</tr>
<tr>
<td>Forced expiratory time &lt; 1 s</td>
<td>123</td>
</tr>
<tr>
<td>Cannot be fitted in groups</td>
<td></td>
</tr>
<tr>
<td>Indian and Pakistani</td>
<td>5,476</td>
</tr>
<tr>
<td>Omani</td>
<td>1,256</td>
</tr>
<tr>
<td>South African</td>
<td>146</td>
</tr>
<tr>
<td>Filipino</td>
<td>316</td>
</tr>
<tr>
<td>Mexico City</td>
<td>4,009</td>
</tr>
<tr>
<td>Iran</td>
<td>6,137</td>
</tr>
<tr>
<td>No permission to publish</td>
<td>3,483</td>
</tr>
<tr>
<td>Outliers</td>
<td>527</td>
</tr>
<tr>
<td>Remaining data</td>
<td>74,187</td>
</tr>
</tbody>
</table>
and log-transformed dependent and explanatory variables, using the Box-Cox-Cole-Green (BCCG) distribution. The optimal degrees of freedom (df) for the spline curve was chosen to minimise the Schwarz Bayesian Criterion (SBC), where adding one df to the model penalises the deviance by $\ln(N)$ units, $N$ being the sample size. As $N$ for males and females ~30,000-40,000, the penalty for an extra df is ~10.3-10.6 deviance units. Thus a parsimonious model with an optimal spline curve was obtained.

Over the 2.5 to 95 year age range the age-specific contribution of the spline in age leads to extensive look-up tables. This is exacerbated if separate equations are derived for different ethnic or geographical groups (henceforth called group). Two simplifying approaches were explored. The first one was to include group (a dummy variable) as a function of age. Thus the general form of the equation was:

$$Y = a + b \cdot H + c \cdot A + \text{spline} + d_1 \cdot \text{group} + d_2 \cdot \text{group} \cdot A$$

where $Y$ = dependent variable, $H$ = standing height (cm); $A$ = age (yr); $a$, $b$, $c$, $d_1$, and $d_2$ are coefficients which vary for each index, and $d_1$ and $d_2$ vary for each group; spline is an age-specific contribution from the spline function. In practice the best fit was obtained after log transformation of the spirometric index (FEV$_1$, FVC, FEV$_1$/FVC ratio), height and age. Group-age interaction did not improve the fit, so that the equation simplified to:

$$Y = a + b \cdot H + c \cdot A + \text{spline} + d_1 \cdot \text{group}$$

The Box-Cox Power Exponential (BCPE) distribution, which also models kurtosis, invariably provided a somewhat better fit than the BCCG distribution. However, the improvement was limited to the tails of the distribution, i.e. at Z-scores beyond $+3$ and $-3$, thus affecting 0.14% of data at either end; we settled for the BCCG distribution, as slightly greater accuracy is clinically meaningless, and implementation of equations simpler.

### 3 SIMPLIFYING LOOK-UP TABLES

As delineated above, a term spline which varies with age arises from fitting a smoothing spline. Please note that age splines were fitted to L, M and S, hence denoted as Mspline, Sspline and Lspline. Figure E1 depicts how spline for predicted FEV$_1$ (hence Mspline) varies with age in healthy white males and females. Particularly in (pre)school children the table for spline needs to be quite detailed. Thus the age-specific table for the 3-95 year age range for L, M and S can each easily have >400 cells for each index (FEV$_1$, etc.). An effort was made to replace each table by an equation, potentially decreasing the complexity of implementing equations in pulmonary function test devices with limited memory dramatically.

It was not possible to satisfactorily fit spline over the entire age range. From age 25 years and above a satisfactory polynomial fit could be obtained.

$$\text{spline} \approx b_0 + b_1 \cdot (A/100) + b_2 \cdot (A/100)^2 + b_3 \cdot (A/100)^3 + b_4 \cdot (A/100)^4 + b_5 \cdot (A/100)^5$$
where A = age in years; seventh degree polynomials were also used to fit Lspline. Coefficients were estimated using linear regression. Thus differences between values predicted using the equation or splines derived by GAMLSS were reduced to a maximum of 0.2%.

Detailed look-up tables are required in children and adolescents, as a few months age difference can affect the predicted values by up to 6%. This is because the predicted values are a power function of age. For example, the linear age coefficient for FEV1 in males is 0.0574; the contribution from the spline at ages 14.75 and 14.0 years is 0.0958 and 0.0684, respectively. Therefore, if one substitutes 14 years into the equation instead of 14.5, the predicted value will be biased by 100·(1 - exp(0.0574·(log(14.5) - log(14)) + 0.0958 - 0.0684)) = -3%.

The look-up tables for the 3-95 years age range, as well as the equations that replace the tables from 25 years up, can be downloaded from www.lungfunction.org/files/lookuptables.xls.

### 4 WORKED EXAMPLES OF CALCULATING PREDICTED VALUES

#### 4.1 Introduction

Predicted values depend on three quantities L, M and S, which are functions of sex, age, height and ethnic group. L measures the skewness, S is the coefficient of variation and M is the predicted value of FEV1, FVC or FEV1/FVC. The lower limit of normal (LLN) and the Z-score are calculated from L, M and S as follows:

\[
\begin{align*}
\text{Predicted value} &= M \\
\text{LLN (5th centile)} &= \exp(\ln(1 - 1.644 \cdot L \cdot S)/L + \ln(M)) \\
\text{Z-score} &= ((\text{measured}/M)^2 - 1)/(L \cdot S) \quad (\text{for } L \neq 0) \\
\% \text{ predicted} &= ((\text{measured}/M) - 100)
\end{align*}
\]

The LMS equations for Caucasians, African Americans, South and North East Asians, valid from 3-95 years, are of this form, with volumes in litres, age in years, and height in cm:

\[
\begin{align*}
L &= q_0 + q_1 \cdot \ln(Age) + Lspline \\
M &= \exp(a_0 + a_1 \cdot \ln(Height) + a_2 \cdot \ln(Age) + a_3 \cdot \text{black} + a_4 \cdot \text{NEA} + a_5 \cdot \text{SEA} + \text{Mspline}) \\
S &= \exp(p_0 + p_1 \cdot \ln(Age) + p_2 \cdot \text{black} + p_3 \cdot \text{NEA} + p_4 \cdot \text{SEA} + \text{Sspline})
\end{align*}
\]

#### 4.2 Example coefficients for FEV1 in males

For the purpose of demonstrating how to use the equations in calculating predicted and derived values, we reproduce tables from the look-up tables for calculating FEV1 in males.

In the case of FEV1 in males, L is a constant, independent of age (Table E4). Table E5 lists the coefficients required to calculate the contribution of splines at specific ages, for subjects between 25-95 years. One should never extrapolate beyond these age ranges!

---

**Table E4 – Regression coefficients for splines for FEV1 in males that may be used to replace look-up tables for ages 25-95 years in systems with limited memory.**

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>S</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>a0</td>
<td>p0</td>
<td>q0</td>
</tr>
<tr>
<td>Height</td>
<td>a1</td>
<td>p1</td>
<td>q1</td>
</tr>
<tr>
<td>Age</td>
<td>a2</td>
<td>p2</td>
<td>q2</td>
</tr>
<tr>
<td>Afr.Am.</td>
<td>a3</td>
<td>p3</td>
<td>q3</td>
</tr>
<tr>
<td>N East Asia</td>
<td>a4</td>
<td>p4</td>
<td>q4</td>
</tr>
<tr>
<td>S East Asia</td>
<td>a5</td>
<td>p5</td>
<td>q5</td>
</tr>
</tbody>
</table>

where

\[
\begin{align*}
\ln() &= \text{natural log transformation} \\
\text{white} &= 1 \text{ if a subject is Caucasian, otherwise } = 0 \\
\text{black} &= 1 \text{ if a subject is African American, otherwise } = 0 \\
\text{NEA} &= 1 \text{ if a subject is North East Asian*, otherwise } = 0 \\
\text{SEA} &= 1 \text{ if a subject is South East Asian*, otherwise } = 0 \\
\end{align*}
\]

Coefficients L0 … q depend on the measurement and sex Mspline, Sspline, Lspline: age-varying coefficients

* Mongoloid people, does not apply to Indian subcontinent.

For 3-95 years:

Linearly interpolate Lspline, Mspline and Sspline from lookup tables as follows:

\[
X_{\text{spline}}(\text{age}) \approx \left[ (\text{age}^2 - \text{age}) \cdot X_{\text{spline}}(\text{age}_1) + (\text{age} - \text{age}_1) \cdot X_{\text{spline}}(\text{age}_2) \right] / (\text{age}_2 - \text{age}_1)
\]

where X represents L, M or S, age = actual age, age1 and age2 represent the ages between which interpolation should be performed.

For 25-95 years one might use polynomial equations (table E5):

\[
\begin{align*}
\text{Mspline} &\approx b_0 + b_1 \cdot (\text{Age} / 100) + b_2 \cdot (\text{Age} / 100)^2 + b_3 \cdot (\text{Age} / 100)^3 + b_4 \cdot (\text{Age} / 100)^4 + b_5 \cdot (\text{Age} / 100)^5 \\
\text{Sspline} &\approx c_0 + c_1 \cdot (\text{Age} / 100) + c_2 \cdot (\text{Age} / 100)^2 + c_3 \cdot (\text{Age} / 100)^3 + c_4 \cdot (\text{Age} / 100)^4 + c_5 \cdot (\text{Age} / 100)^5 \\
\text{Lspline} &\approx d_0 + d_1 \cdot (\text{Age} / 100) + d_2 \cdot (\text{Age} / 100)^2 + d_3 \cdot (\text{Age} / 100)^3 + d_4 \cdot (\text{Age} / 100)^4 + d_5 \cdot (\text{Age} / 100)^5 + d_6 \cdot (\text{Age} / 100)^6 + d_7 \cdot (\text{Age} / 100)^7
\end{align*}
\]

---

**Figure E1 – The contribution of the age-spline to predicted FEV1 in males and females varies with age.**
Table E5 – Regression coefficients for splines for FEV₁ in males that may be used to replace look-up tables for ages 25-95 years in systems with limited memory.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Mspline</th>
<th>Sspline</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>b₀</td>
<td>c₀</td>
<td>d₀</td>
</tr>
<tr>
<td>Age</td>
<td>b₁</td>
<td>c₁</td>
<td>d₁</td>
</tr>
<tr>
<td>Age₂</td>
<td>b₂</td>
<td>c₂</td>
<td>d₂</td>
</tr>
<tr>
<td>Age₃</td>
<td>b₃</td>
<td>c₃</td>
<td>d₃</td>
</tr>
<tr>
<td>Age₄</td>
<td>b₄</td>
<td>c₄</td>
<td>d₄</td>
</tr>
<tr>
<td>Age₅</td>
<td>b₅</td>
<td>c₅</td>
<td>d₅</td>
</tr>
<tr>
<td>Age₆</td>
<td>b₆</td>
<td>c₆</td>
<td>d₆</td>
</tr>
<tr>
<td>Age₇</td>
<td>b₇</td>
<td>c₇</td>
<td>d₇</td>
</tr>
</tbody>
</table>

4.3 Five worked examples

4.3.1 White boy

We wish to calculate the predicted FEV₁, % predicted, 5th centile LLN and Z-score for a white boy age 4.8 yr, height 107 cm, FEV₁ = 0.800 L.

For linear interpolation for age 4.8 yr we consult the look-up table:

<table>
<thead>
<tr>
<th>Age (yr)</th>
<th>Mspline</th>
<th>Sspline</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.75</td>
<td>-0.0769</td>
<td>0.1592</td>
</tr>
<tr>
<td>5</td>
<td>-0.0752</td>
<td>0.1535</td>
</tr>
</tbody>
</table>

Mspline = -0.0769 + ((4.8-4.75)/0.25)(-0.0752 - (-0.0769)) = -0.0766
Sspline = 0.1592 + ((4.8-4.75)/0.25)(0.1535 - 0.1592) = 0.1581

M = FEVₐₚrₑ𝑐ₑ𝑑 = exp(-10.3420 + 2.2196·ln(107) + 0.0574·ln(48) - 0.0766) = 1.0442 L
% predicted = (0.800/1.0442)·100 = 76.6%
L = 0.8866 + 0.0850·ln(4.8) = 1.0199
S = exp(-2.3268 + 0.0798·ln(4.8) + 0.0327 - 0.0007) = 0.1395
Z-score = ((measured/predicted)¹ - 1)/(L·S) = -0.0003

We can also replace the look-up tables by the equations in Table E5:

Mspline = -0.0404
Sspline = 0.0003
M = FEVₐₚₑᶜₑᵈ = exp(-10.3420 + 2.2196·ln(175) + 0.0574·ln(53) - 0.0881 - 0.0404) = 3.3911 L
% predicted = (2.410/3.3911)·100 = 71.1%
L = 0.8866 + 0.0850·ln(53) = 1.2241
S = exp(-2.3268 + 0.0798·ln(53) + 0.0327 + 0.0003) = 0.1395
Z-score = ((measured/predicted)¹ - 1)/(L·S) = -2.00

4.3.2 African American boy

We wish to calculate the predicted FEV₁, % predicted, 5th centile LLN and Z-score for an African American boy age 12.2 yr, height 165 cm, FEV₁ = 2.210 L.

For linear interpolation to age 12.2 yr we consult the look-up table:

<table>
<thead>
<tr>
<th>Age (yr)</th>
<th>Mspline</th>
<th>Sspline</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>-0.0176</td>
<td>-0.0387</td>
</tr>
<tr>
<td>12.25</td>
<td>-0.0101</td>
<td>-0.0339</td>
</tr>
</tbody>
</table>

Mspline = -0.0176 + ((12.2-12)/0.25)(-0.0101 - (-0.0176)) = -0.0116
Sspline = -0.0387 + ((12.2-12)/0.25)(-0.0339 - (-0.0387)) = -0.0349
M = FEVₐₚₑᶜₑᵈ = exp(-10.3420 + 2.2196·ln(152) + 0.0574·ln(12.2) - 0.1589 - 0.0116) = 2.1860 L
% predicted = (2.405/2.1860)·100 = 110.0%

4.3.3 South East Asian adult male

We wish to calculate the predicted FEV₁, % predicted, 5th centile LLN and Z-score for a South East Asian male age 53 yr, height 175 cm, FEV₁ = 2.410 L.

Calculations can be performed using the look-up table for L, M, and S, or by replacing the latter with the equations in Table E5. First using the look-up tables:

<table>
<thead>
<tr>
<th>Age (yr)</th>
<th>Mspline</th>
<th>Sspline</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>-0.0404</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

Mspline = -0.0404
Sspline = 0.0003
M = FEVₐₚₑᶜₑᵈ = exp(-10.3420 + 2.2196·ln(175) + 0.0574·ln(53) - 0.0881 - 0.0404) = 3.3911 L
% predicted = (2.410/3.3911)·100 = 71.2%
L = 0.8866 + 0.0850·ln(53) = 1.2241
S = exp(-2.3268 + 0.0798·ln(53) + 0.0327 + 0.0003) = 0.1395
Z-score = ((measured/predicted)¹ - 1)/(L·S) = -2.00

We can also replace the look-up tables by the equations in Table E5:

Mspline = 0.3901 - 1.0579·(53/100) + 1.4743·(53/100)² - 68.1649·(53/100)³ + 127.1964·(53/100)⁴ - 109.6777·(53/100)⁵ + 35.6832·(53/100)⁶ = -0.0427
Sspline = 0.8866 + 0.0850·ln(53) = 1.0199
L = 0.8866 + 0.0850·ln(53) = 1.2241
S = exp(-2.3268 + 0.0798·ln(53) + 0.0327 + 0.0003) = 0.1395
Z-score = ((measured/predicted)¹ - 1)/(L·S) = -2.00

4.3.4 South East Asian adult female

We wish to calculate the predicted FEV₁, % predicted, 5th centile LLN and Z-score for a South East Asian female age 39.1 yr, height 165 cm, FEV₁ = 2.210 L.

Calculations can be performed using the look-up table for L, M, and S, or by replacing the latter with the equations in Table E5.

<table>
<thead>
<tr>
<th>Age (yr)</th>
<th>Mspline</th>
<th>Sspline</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>-0.0404</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

Mspline = -0.0404
Sspline = 0.0003
M = FEVₐₚₑᶜₑᵈ = exp(-10.3420 + 2.2196·ln(175) + 0.0574·ln(53) - 0.0881 - 0.0404) = 3.3911 L
% predicted = (2.410/3.3911)·100 = 71.2%
L = 0.8866 + 0.0850·ln(53) = 1.2241
S = exp(-2.3268 + 0.0798·ln(53) + 0.0327 + 0.0003) = 0.1395
Z-score = ((measured/predicted)¹ - 1)/(L·S) = -2.00
Online Supplement: Multi-ethnic reference values for spirometry: the Global Lung Function 2012 Equations

Table E6 – Regression coefficients for FEV1 for calculating M (predicted value), S (coefficient of variation) and L (skewness). The contributions of splines must be added to the calculated values; they are available in look-up tables.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Mspline</th>
<th>Sspline</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>a0</td>
<td>-9.6987</td>
<td>p0</td>
</tr>
<tr>
<td>Height</td>
<td>a1</td>
<td>2.1211</td>
<td>p1</td>
</tr>
<tr>
<td>Age</td>
<td>a2</td>
<td>-0.0270</td>
<td>p2</td>
</tr>
<tr>
<td>Afr.Am.</td>
<td>a3</td>
<td>-0.1484</td>
<td>p3</td>
</tr>
<tr>
<td>N East Asia</td>
<td>a4</td>
<td>-0.0149</td>
<td>p4</td>
</tr>
<tr>
<td>S East Asia</td>
<td>a5</td>
<td>-0.1206</td>
<td>p5</td>
</tr>
</tbody>
</table>

Mspline = 0.1112 + (0.1/0.25)·(0.1097 - 0.1112) = 0.1105

Sspline = -0.0919 + (0.1/0.25)·(-0.0913 - (-0.0919)) = -0.0917

Lspline = 0

M = FEV1, predicted = \exp(-9.6987 + 2.1211·\ln(165) - 0.0270·\ln(39.1) - 0.1206 + 0.1105) = 2.7800 L

% predicted = (2.210/2.7800)·100 = 79.5 %

S = \exp(-2.3765 + 0.0972·\ln(39.1) + 0.0733 - 0.0917) = 0.1302

L = 1.1540

LLN = \exp(\ln(1 - 1.644·L·S)/L + \ln(M))
    = \exp(\ln(1 - 1.644·1.1540·0.1302)/1.1540 + \ln(2.7800))
    = 2.1741 L

Z-score = ((measured/predicted)L - 1)/(L·S)
         = ((2.210/2.7800)1.1540 - 1)/(1.1540·0.1302) = -1.55

Table E7 – Regression coefficients for splines for FEV1 in females that may be used to replace look-up tables for ages 25-95 years in systems with limited memory.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Mspline</th>
<th>Sspline</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>b0</td>
<td>0.0552</td>
<td>c0</td>
</tr>
<tr>
<td>Age</td>
<td>b1</td>
<td>1.6032</td>
<td>c1</td>
</tr>
<tr>
<td>Age2</td>
<td>b2</td>
<td>-6.4855</td>
<td>c2</td>
</tr>
<tr>
<td>Age3</td>
<td>b3</td>
<td>10.2741</td>
<td>c3</td>
</tr>
<tr>
<td>Age4</td>
<td>b4</td>
<td>-9.8646</td>
<td>c4</td>
</tr>
<tr>
<td>Age5</td>
<td>b5</td>
<td>3.8808</td>
<td>c5</td>
</tr>
</tbody>
</table>

We can also replace the look-up tables by equations (Table E7):

Mspline = 0.0552 + 1.6032·(39.1/100)·6.4855·(39.1/100)2 + 10.2741·(39.1/100)·29.4400·(39.1/100)3 - 9.8646·(39.1/100)4 + 3.8808·(39.1/100)5 = 2.7800 L

% predicted = (2.210/2.7775)·100 = 79.6 %

S = \exp(-2.3765 + 0.0972·\ln(39.1) + 0.0733 - 0.0894) = 0.1305

L = 1.1540

LLN = \exp(\ln(1 - 1.644·L·S)/L + \ln(M))
    = \exp(\ln(1 - 1.644·1.1540·0.1305)/1.1540 + \ln(2.7775))
    = 2.1707 L

Z-score = ((measured/predicted)L - 1)/(L·S)
         = ((2.210/2.7775)1.1540 - 1)/(1.1540·0.1305) = -1.54

Figure E2 – Standing height as a function of age in groups from various parts of the world. Caucasians and African Americans are up to 10 cm taller than other ethnic groups.

N East Asia = North East Asia; Afr.Am. = African American; Mx.Am. = Mexican American; Lat.Am. = Latin American; India+Pak = India and Pakistan; S East Asia = South East Asia; N.Afr.+Iran = Algeria, Tunisia and Iran.
As demonstrated above, there are slight differences in the results calculating values using the look-up table or the equations replacing them. The discrepancies are not clinically important; nevertheless use of look-up tables and linear interpolation is the recommended procedure.

5 DIFFERENCES IN STANDING HEIGHT BETWEEN GROUPS

Stature, the main determinant of pulmonary function, differed significantly between populations (Figure E2). The coefficient of variation (CoV) for height is largest in preschool children (Figure E3), declining rapidly towards adolescence followed by an increase, reflecting differences in the age of onset and end of the pubertal growth spurt. There is then a drop until about age 30 years, followed by a small but steady increase towards old age. With one exception, the maximum difference between groups is about 0.01 (1%). The CoV is considerably larger in Indian and Pakistani schoolchildren and adolescents than in others (Figure E3), probably due to differences in socio-economic conditions; differences in stature and CoV between Pakistani and Indian children were small.

6 DATA

6.1 Latin-American data

Five datasets (Mexico City, Sao Paulo, Caracas, Montevideo, Santiago) [3] related to adults, one to Mexican children and young adults [4]. There were pronounced differences in height for age between centres, in males ranging from 12 cm at age 45 to 7 cm at age 80 years (Figure E4), people in Mexico City being smallest. Predicted values for spirometry derived from each individual dataset also produced appreciable differences (Figure E5).

6.2 East Asian data

Nine datasets were available from Hong Kong [5,6], Taiwan [7], Thailand [8], the USA [9], Korea [10], China [11] (data...
not used for present analyses, see printed text), North China [12], and a dataset on children aged 3-6 years [13]. Regression analysis revealed significant differences for spirometric indices between centres. Using GAMLSS, a best fit of height for age was derived from the collated data of East Asians; calculated height for age was then used to derive predicted values for spirometric indices for each centre, allowing comparison between centres that was unaffected by differences in height. Predicted values for FEV₁ and FVC from mainland China (unpublished data from North and South China collected in 2008), North China [12] and Korea [10] came out systematically higher than in the remaining 6 datasets (data collected between 1996-2002 in Hong Kong, Taiwan, Thailand, USA and China), which agreed remarkably well (Figure E6). No evidence was found that this related to methodological differences, or to unrepresentative samples arising from small sample size. There were significant differences in standing height between centres, Chinese in the USA and the North of China (i.e. north of the Huai River and Qinling Mountains) being tallest, and people in Thailand shortest: at age 50 yr maximum differences were 5.4 cm in females, and 5.7 cm in males. Data collection spanned 18 years (1990-2008), during which time changes occurred in Oriental societies which significantly affected the availability of food and education. Better health conditions are associated with an increase in standing height, mainly due to an increase in leg length [14], although the latter finding could not be reproduced recently [15]. Average height for someone aged for example 40 yr in 1990 may differ from that of a 40 yr old person in 2008. Displaying standing height as a function of birth date, rather than calendar age, removes temporal differences, allowing to better depict differences between populations (Figure E7). The slopes are steepest for males and females in the USA and Korea, and flattest for Taiwan, leading to curves crossing over at about the year 1960 (Figure E7). These slopes reflect a combination of diminishing height with advancing age and secular trends in height. For example, for mainland China the estimated rise in standing height between 1979-1995 for a 17 years old

![Figure E6](image1.png)  
Figure E6 – Predicted FEV₁ and FVC in 9 different datasets from East Asian subjects. Data from South East Asia are systematically shorter than those from North East Asia.

![Figure E7](image2.png)  
Figure E7 – Relationship between year of birth and standing height in adult East Asian subjects born in Hong Kong, China, Thailand, Taiwan, the USA and Korea. The slope of the line reflects both a secular trend and a change in height with ageing.
Figure E8 – Coefficient of variation (CoV) of FEF_{25-75} and FEF_{75} in males and females. Note the very large variability, particularly in adults, which severely limits the use of these indices for diagnostic purposes. However, this does not preclude a more favourable coefficient of variation for within-subject comparisons, nor the use of these indices in etiological studies where differences between groups may provide valuable clues.

NE Asia = North East Asia; SE Asia = South East Asia, Afr.Am. = African American.

Figure E9 – Comparison of predicted values for FVC (right panels) and FEV₁ (left panels) in males and females. GLI = Global Lungs Initiative (present study), Stanojevic [27], NHANES III [22], ECSC = European Community for Steel and Coal [26], HSE = Health Survey for England 1995-1996 [23], LuftiBus [24], SAPALDIA [25], Polgar [28], Zapletal [29]. Predicted values from the ECSC in adults, from the LuftiBus study in young adults, and from Polgar and Zapletal in children and adolescents, do not agree well with the others. Data from Stanojevic and the GLI practically overlap. Note the different scaling from 0-20 years and 20-100 years, so as to show greater detail in the paediatric age range. Graphs were generated using mean height for age in Caucasians, hence the above differences in predicted values cannot be accounted for by differences in height.
person was 1.0 and 2.0 cm per decade for urban and rural males, respectively, and 0.5 and 0.7 cm per decade for urban and rural females, respectively [16]. At the same time the coefficient of variation for stature widened, indicating greater regional variation in height for age [16,17]. However, the well-known average height difference between southern and northern Chinese for urban populations has narrowed from 6-7 cm in the 1920s to 2-3 cm in the 1990s [17]. Figure E7 shows that the subjects from mainland China and Korea were on average taller, yet shorter than Orientals living in the USA. Differences in pulmonary function remaining after standardising for the same height for age (Figure E6) might therefore reflect differences in body build. This is in keeping with various reports [15-21]. The scale short of differences in FEV₁ and FVC, and the limited time span between data collections are not compatible with a secular trend.

6.3 Handling of longitudinal datasets

Two longitudinal studies of schoolchildren [18] and adolescents [19] were transformed into a cross-section by selecting at random one record from a person’s available measurements so that the new cross-sectional data set had a representative age distribution. In another follow-up study of school children [20] only the data collected at the first occasion were selected. In a longitudinal study of adults [21] the third record was selected.

7 FEF₂₅₋₇₅ AND FEF₇₅

The coefficients of variation of FEF₂₅₋₇₅ and FEF₇₅ rise to very high values in adults (Figure E8).

8 COMPARISON OF PREDICTED VALUES

The present set of prediction equations, using collated data, was compared with those from four recent large datasets [22-25]. Regression equations were derived using GAMLSS [1] in R [Version 2.14.1; R Foundation, http://www.r-project.org]. Predicted values and their lower limits of normal were calculated as a function of age, using the mean height for age in Caucasians; the latter was obtained by regressing height on age, using GAMLSS. Predicted values according to the ECSC/ERS [26], widely used in Europe, and predicted values from Stanoevic [27] which are being used increasingly, were added to the comparison (see also Table 5 in printed paper), as were those from Polgar and Zapletal [28,29] (Figure E9). Except for the ECSC/ERS predicted values and those from Polgar and Zapletal, which differ systematically from the other ones, differences are marginal. Hence, the use of collated data has not led to loss of accuracy and precision.

Acknowledgement

The authors are extremely grateful to all individuals and organisations who contributed data and information to the Global Lungs Initiative. Without their help, contributions and mutual trust this project would have been impossible. The extensive statistical review by C. Schindler is also gratefully acknowledged. This study is based on data from 70 centres, including the MESA study. The MESA and MESA Lung Studies are conducted and supported by the National Heart, Lung and Blood Institute (NHLBI) in collaboration with the MESA and MESA Lung Investigators. In addition to review by representatives of all bodies contributing data to the GLI, this manuscript has been reviewed by the MESA investigators for scientific content and consistency of data interpretation with previous MESA publications and significant comments incorporated prior to submission for publication. A full list of participating MESA Investigators and institutions can be found at http://www.mesa-nhlbi.org/.

References

9. MESA Lung Study, courtesy dr G. Barr.


