



ERS TASK FORCE

Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations

Philip H. Quanjer, Sanja Stanojevic, Tim J. Cole, Xaver Baur, Graham L. Hall, Bruce H. Culver, Paul L. Enright, John L. Hankinson, Mary S.M. Ip, Jinping Zheng, Janet Stocks and the ERS Global Lung Function Initiative

ABSTRACT: The aim of the Task Force was to derive continuous prediction equations and their lower limits of normal for spirometric indices, which are applicable globally. Over 160,000 data points from 72 centres in 33 countries were shared with the European Respiratory Society Global Lung Function Initiative. Eliminating data that could not be used (mostly missing ethnic group, some outliers) left 97,759 records of healthy nonsmokers (55.3% females) aged 2.5–95 yrs.

Lung function data were collated and prediction equations derived using the LMS method, which allows simultaneous modelling of the mean (μ), the coefficient of variation (σ) and skewness (λ) of a distribution family.

After discarding 23,572 records, mostly because they could not be combined with other ethnic or geographic groups, reference equations were derived for healthy individuals aged 3–95 yrs for Caucasians ($n=57,395$), African-Americans ($n=3,545$), and North ($n=4,992$) and South East Asians ($n=8,255$). Forced expiratory value in 1 s (FEV₁) and forced vital capacity (FVC) between ethnic groups differed proportionally from that in Caucasians, such that FEV₁/FVC remained virtually independent of ethnic group. For individuals not represented by these four groups, or of mixed ethnic origins, a composite equation taken as the average of the above equations is provided to facilitate interpretation until a more appropriate solution is developed.

Spirometric prediction equations for the 3–95-age range are now available that include appropriate age-dependent lower limits of normal. They can be applied globally to different ethnic groups. Additional data from the Indian subcontinent and Arabic, Polynesian and Latin American countries, as well as Africa will further improve these equations in the future.

KEYWORDS: Lower limit of normal, predicted values, spirometry, statistical modelling, z-score

CONTENTS

Introduction	1325
Objectives	1326
Methods	1326
Recruitment and data collection	1326
Data	1326
Quality control	1326
Indices considered	1326
Statistical analyses	1326
Prediction models	1326
Defining groups	1327

AFFILIATIONS

For a full list of the authors' affiliations, see the Acknowledgements.

CORRESPONDENCE

P.H. Quanjer
Dept of Pulmonary Diseases and
Dept of Paediatrics
Erasmus Medical Centre
Erasmus University
Rotterdam
The Netherlands
E-mail: pquanjer@xs4all.nl

Received:

May 22 2012

Accepted after revision:

June 06 2012

First published online:

June 27 2012

Results	1327
Data	1327
Stature	1327
Spirometry datasets by region	1328
Latin America	1328
East Asia	1328
African–American	1328
Mexican–American	1328
North Africa and Iran	1328
Indian subcontinent	1329
Oman	1329
Japan	1329
Pooling groups	1329
Other	1329
Prediction models	1330
Simplifying look-up tables	1330

Proportional differences between ethnic groups	1330
Effect of collation on predicted values and LLN	1330
Age-related change in pulmonary function	1330
Entering height and age into the equations	1330
Discussion	1332
Representativeness of equations for various groups ..	1334
Data quality	1336
Secular trends	1336
Age-related trends in FEV ₁ and FVC	1336
Mixed ethnic descent	1336
Application in clinical practice	1336
Clinical decision making	1337
Strengths and weaknesses	1338
Future developments	1338
Conclusions	1338
Recommendations	1338

INTRODUCTION

Pulmonary function tests fulfil a pivotal role in respiratory medicine. They are used to diagnose airways obstruction, assess its severity and prognosis, delineate risk factors (*e.g.* pre-operative assessment), detect early lung disease, and monitor for normal lung growth and lung function decline. Unlike the majority of biological indices in medicine, such as plasma concentrations of chemical analytes or hormones, pulmonary function varies with age, standing height, sex and ethnicity. Therefore, test results need to be compared to predicted values, and lower and upper limits of normal (LLN and ULN, respectively) that are appropriate for the individual being tested. There is a plethora of published reference equations [1], mostly for spirometric indices, and most publications relate to Caucasians (we will use this term to denote people of European ancestry, in line with the National Institutes of Health [2]). With relatively few exceptions, the appropriateness of the chosen model was not tested and the LLN or ULN were not properly derived. Also many prediction equations are based on small numbers of subjects, using data collected decades ago so that changes in spirometric methodology and secular trends (*i.e.* a trend in pulmonary function in successive birth cohorts) may affect the applicability to present day measurements. Few equations take into account the changing relationship between lung function and height during the adolescent growth spurt [3–6]. Almost invariably prediction equations cover a limited age range, such as childhood, school age or adulthood, leading to discontinuities as individuals move from one set of equations to the next. For example, based on predicted values from POLGAR and PROMADHAT [7] up to 18 yrs of age and from the European Community for Steel and Coal (ECSC)/European Respiratory Society (ERS) [8] for adults, a 175-cm tall young male producing a forced vital capacity (FVC) of 4.2 L is at 97% of predicted at 17.9 yrs of age, but at 83% on his 18th birthday. Such discrepancies are even more marked in individuals who are short for their age, a frequent sequel of childhood chronic lung disease. Many laboratories accept the default settings for predicted values for different age ranges offered by the manufacturer and are insufficiently aware of these problems.

Thus, there is a need for prediction equations based on a sufficiently large representative population sample across the entire age range, using up-to-date methodology. The collection of large numbers of lung function test results in lifelong healthy nonsmokers is, however, time consuming and costly. It has been shown that collating data collected in different centres using state-of-the-art techniques and appropriate quality control is a valid and very cost-effective way of deriving prediction equations [9, 10]. Another major breakthrough was the application of a novel statistical technique (GAMLSS; www.lungfunction.org/files/GAMLSS-in-action.zip) [11] to collated data from 11 countries; this allowed modelling of spirometric indices from early childhood to old age, and also produced age-specific values for the LLN [12, 13]. In a study comprising 43,032 Caucasians from 30 centres, differences in the level of spirometric indices and their variability were shown to be related to sample size, with no evidence of secular trends over a 30-yr period [14], validating the use of collated data. Another benefit of using collated data was the identification of a biphasic trend in the forced expiratory volume in 1 s (FEV₁)/FVC ratio in childhood and adolescence [15].

In 2005, the American Thoracic Society (ATS) and the ERS published recommendations for standardised lung function testing [16]. A set of spirometric reference equations for adults [17] and a set for children and adolescents [6] were recommended for use in the USA; this covered Caucasians, African- and Mexican-Americans. The lack of recommendations for the rest of the world underlines the urgent need to derive all-age reference equations that are valid worldwide, applicable to as many ethnic groups as possible. From 2006, an increasing number of centres agreed to share data with one of the authors (P.H. Quanjer); subsequently the Global Lung Function Initiative (GLI) was established in Berlin in September 2008, acquiring ERS Task Force status in April 2010. The GLI was subsequently endorsed by the ATS, Australian and New Zealand Society of Respiratory Science (ANZSRS), Asian Pacific Society for Respirology (APSR), the Thoracic Society of Australia and New Zealand (TSANZ) and the American College of Chest Physicians (ACCP).

Objectives

The objectives of the GLI are to establish improved international spirometry reference equations that: 1) are based on individual lung function data collected under standardised measurement conditions with documented equipment and software; 2) are modelled using modern statistical techniques to allow continuous equations across the entire age range from early childhood to old age; 3) allow flexible and appropriate methods of interpretation using limits of normality, which adjust for the heterogeneity of between-subject variability according to sex, ethnic group, age and lung function parameters; 4) are clinically useful and can be incorporated into commercially available equipment; and 5) are reported in such a manner as to give a clear indication of where the subject lies with respect to the “normal range”.

METHODS

Recruitment and data collection

We contacted individuals, groups or organisations that were known to have collected representative spirometry data from asymptomatic lifelong nonsmokers using methods and techniques that complied with temporal international recommendations. Most invitations arose from searches of the literature for titles in peer-reviewed journals. All data were anonymised prior to submission, and accompanied by information detailed in a data collection template. Any person, group or organisation sharing data with the working group specified in writing that their ethics committees or organisations had given permission for the data to be used in a research publication.

Data

Datasets were obtained from 73 centres (n=160,330 initially). In France it is prohibited by law to record ethnicity; as a result 63,031 records known to be of mixed ethnic population could not be included in the final analyses. In the remaining datasets, ethnicity could not be traced in an additional 834 cases. In 805 cases data were discarded because they comprised subjects with suspected asthma. In 123 cases data could not be used to derive reference values because forced expiratory time was <1 s. Records with transcription errors that could not be resolved, with missing values for sex, age, height, FEV₁ or FVC, or where the FEV₁/FVC ratio was >1.0 were discarded. Since virtually all data had been previously used in publications, there were very few errors. Datasets from India, Pakistan, Iran, Oman, the Philippines and South Africa were either too small in number for analysis, or could not be combined into groups with other sets (n=17,341). One dataset (n=3,483) could not be used until the data had first been published by the authors. As the statistical analyses are sensitive to outliers, in subsequent analyses data points that yielded a z-score < -5.0 or >5.0 were identified as outliers (n=526) and excluded from further analyses. This left data on 31,856 males and 42,331 females aged 2.5–95 yrs (fig. 1 and tables E1, E2 and E3 in the online supplementary material).

Quality control

Contributors were required to meet certain conditions [18], and data were only accepted if internationally agreed standards had been applied at the time of data collection. The large number of participating centres and the very large number of

data precluded rigorous *post hoc* quality control of all original spirometry.

Indices considered

This study is limited to spirometric indices, analysis of data on lung volumes and transfer factor being deferred to a later stage. Prediction equations were derived for the FEV₁, FVC and FEV₁/FVC across the entire age range. For children aged 3–7 yrs (an age range chosen because the forced expiratory time usually exceeds 1 s in older children), the FEV_{0.75} and FEV_{0.75}/FVC were also derived. Data on FEV_{0.75}, FEV_{0.75}/FVC and forced expired flow at 75% of FVC (FEF_{75%}) has been exhaled were available only for Caucasians. Data (n=36,831) on FEF_{25–75%} were available in 21 datasets. As very few data became available on FEV_{0.5} this index was not analysed.

Statistical analyses

Datasets can be characterised in terms of the following distributional details: the mean value (M), the coefficient of variation (scatter; S) and an index of skewness (location; L), in short LMS. Analyses were performed with the LMS (lambda, mu and sigma) method using the GAMLSS package [11] in the statistical software R (version 2.14.1; www.r-project.org); this allows modelling of each component of the distribution. GAMLSS (version 4.1–2) was used to derive the best fitting function of each outcome as a function of age and height in males and females. The statistical methods used have been described previously [19].

Prediction models

The LMS method, imbedded in GAMLSS, allows modelling the expected mean, coefficient of variation (CoV) and skewness. In addition, complex effects of explanatory variables on the dependent variable can be modelled using splines, which allow the dependent variable to vary smoothly (non-linearly) as a function of an explanatory variable. Thus, a continuous, smooth fit over the entire age range can be obtained by the use of splines.

Applying the methodology described by COLE *et al.* [19] the best fit was estimated using the Box-Cox-Cole-Green distribution. The optimal degrees of freedom (df) for the spline curve was chosen to minimise the Schwarz Bayesian Criterion, where adding 1 df to the model penalises the deviance by ln(N) units, where N is the sample size. As N for males and females was ~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 for males and females. The general form of the equation was:

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

where Y is the dependent variable, H is the standing height (cm), A is age (yrs), and a, b, c, d₁ and d₂ are coefficients which vary for each ethnic group, and spline is an age-specific contribution from the spline function. Group is a dummy variable with values of 0 or 1 indicating ethnicity, where Caucasians are the reference. Any of Y, H or A may be log transformed (refer to online supplementary material). Essentially, we are therefore dealing with a linear regression equation with an age-specific correction in the form of the age-spline.

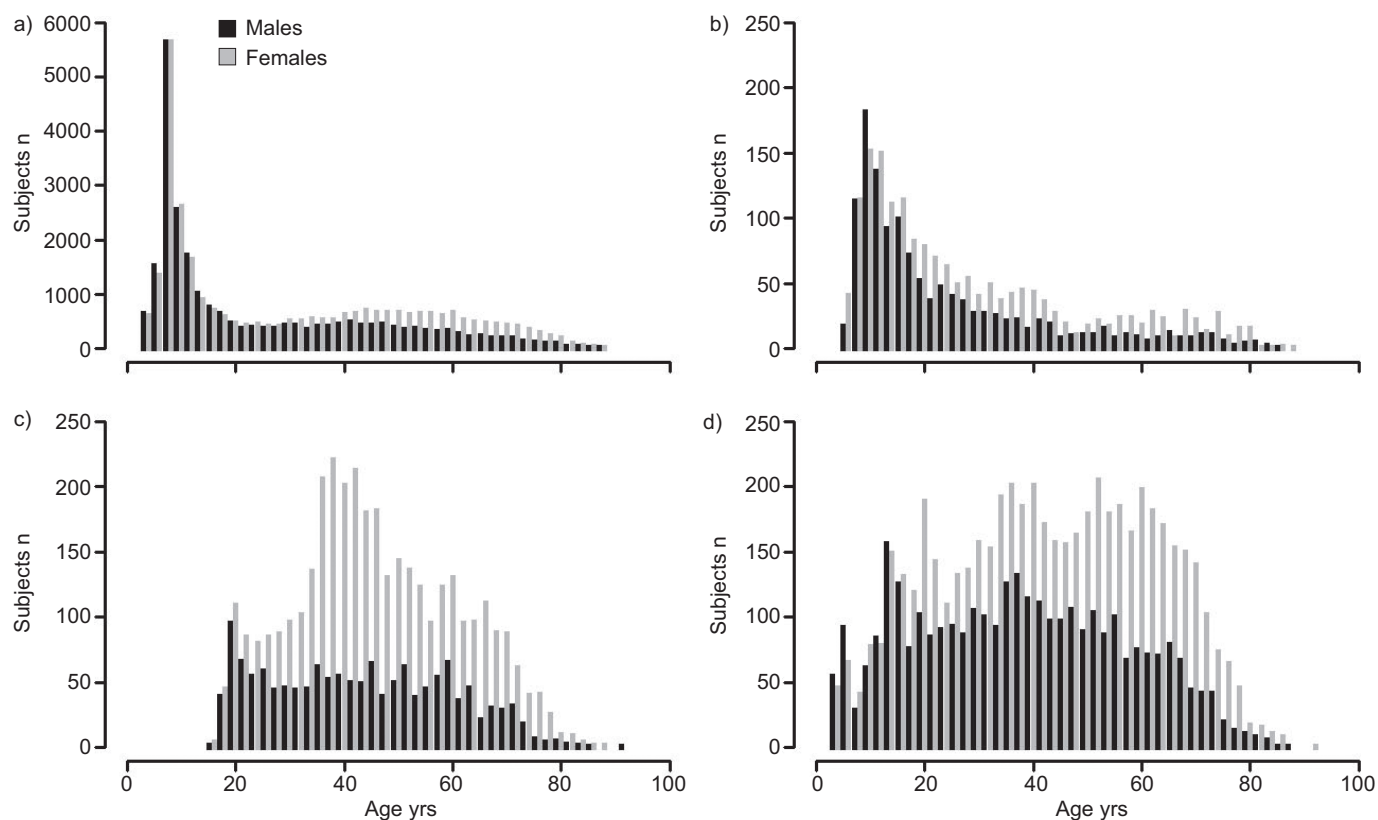


FIGURE 1. Age distribution in the final sample ($n=74,187$) of a) Caucasians, b) African-Americans, c) North East Asians and d) South East Asians. The large number of 8-yr-olds did not lead to any bias in predicted values [14].

The goodness of fit was judged from inspection of normal Q-Q plots, worm plots [20], the distribution of residuals as a function of age and predicted value, and from density plots of residuals.

When collating data covering different age ranges, there is a risk of one or two outlying datasets distorting the general trend, which GAMLSS might then incorporate into the smoothing spline. However, inspecting the age distribution of residuals of all datasets confirmed the absence of such outliers.

Defining groups

Data were available from five continents, comprising different ethnic groups. In datasets with mixed ethnic groups, ethnicity had been coded for each individual. In all other cases data allocation was based on the country of origin of the dataset. Preliminary analyses showed that certain datasets fit together. Hence, the following groups were formed: Caucasian (persons having origins in any of the original peoples of Europe, the Middle East or North Africa); African-American; Mexican-American; Indian subcontinent; North Africa and Iran; East Asia; Latin America; and Oman. Regression analysis was performed within each group; the residuals were then displayed as a function of age and inspected for offsets and variability. Individual datasets generally differed from the overall mean with respect to both offset and between-subject variability. If considerable disparities with other datasets in a group were found, the centre was contacted with a view to

clarifying potential causes for differences. Groups were only included in the final multi-ethnic model if there was fair agreement between their individual datasets in predicted mean and coefficient of variation.

RESULTS

Data

The total number of subjects originally included in the study (*i.e.* prior to exclusion for reasons listed in the Data section) was 97,759 (55.3% females), age range 2.5–95 yrs (table 1 and tables E2 and E3 in the online supplementary material); 47.7% were aged <20 yrs and 0.8% were aged >80 yrs. These data were combined from 72 datasets from 33 countries (table E1 in the online supplementary material). Representation was poor from South America and absent from Malaysia, Indonesia and sub-Saharan Africa.

Stature

Stature is the main determinant of pulmonary function; therefore, we investigated whether stature differed significantly between populations (figs E2 and E3 in the online supplementary material). While there were significant differences in stature for age between populations, the variability between groups was minimal. The within-group CoV was largest in preschool children (fig. E3 in the online supplementary material) and declined rapidly towards adolescence, which was followed by an increase reflecting differences in the timing of the pubertal growth spurt. There was then a drop until about 30 yrs of age, followed by a small but steady

increase towards old age. With one exception, the maximum difference in the CoV between populations was ~1%. The CoV was considerably larger in Indian and Pakistani school children and adolescents than in other populations (fig. E3 in the online supplementary material).

Spirometry datasets by region

Data were available from the following countries: Algeria, Australia, Austria, Brazil, Canada, Chile, China, France, Germany, Iceland, India, Iran, Israel, Italy, Mexico, the Netherlands, Norway, Oman, Pakistan, Philippines, Poland, Portugal, South Africa, South Korea, Sweden, Switzerland, Taiwan, Thailand, Tunisia, UK, USA, Uruguay and Venezuela. Not all of these data could be used (table E1 in the online supplementary material).

Latin America

Six datasets were available from Latin America: five related to adults (Mexico City, São Paulo, Caracas, Montevideo and Santiago) [10], and the remaining one to Mexican children and young adults [21]. The samples differed in height (fig. E4 in the online supplementary material) and predicted spirometric values (fig. E5 in the online supplementary material). These differences could not be explained by altitude or location, and were probably due to sampling variability (since all but one of the datasets was of limited size [14]). Furthermore, no data were available between 25 and 40 yrs of age.

East Asia

Nine datasets with 13,247 records (table E3 in the online supplementary material) were available from Hong Kong (China) [22, 23], Taiwan [24], Thailand [25], the USA (personal communication; the Multi-Ethnic Study of Atherosclerosis (MESA) investigators R.G. Barr, P.L. Enright and J.L.

Hankinson), Korea [26], China [27] (personal communication; multicentre data from mainland China, courtesy of J. Zheng), and one from Chinese children aged 3–6 yrs [28]; all datasets were from urban populations. There were significant differences in standing height between centres, with Chinese people in the USA and mainland China being taller, and people in Thailand, Taiwan, Hong Kong and southwest China being shorter (figs E6 and E7 in the online supplementary material).

Regression analysis revealed significant differences for spirometric indices between centres. While there was remarkable agreement between six of the datasets (data collected between 1996–2002 in Hong Kong, Taiwan, Thailand, USA and China), predicted values for FEV₁ and FVC were significantly lower than those derived from the remaining two datasets, collected in North China [27] and Korea [26] (fig. E6 in the online supplementary material). No evidence was found that this related to methodological differences or to unrepresentative samples arising from small sample size. The scale of differences and the limited time span between data collections are not compatible with a secular trend. In view of this, separate predicted values were derived for East Asians from the North and the South.

African-American

Four African-American datasets were available [6, 17, 29] (personal communication; the MESA investigators R.G. Barr, P.L. Enright and J.L. Hankinson) comprising 1,520 males and 2,025 females. Since two sets had limited numbers and age ranges, data were pooled and residuals derived for each centre. The differences between these datasets were within the range compatible with sampling variability. Differences in the FEV₁/FVC ratio between centres were trivial (maximum z-score difference 0.06), signifying that deviations in FEV₁ and FVC from the overall mean were proportional.

Mexican-American

Four Mexican-American datasets were available [6, 17, 29] (personal communication; G. Barr) comprising 1,622 males and 2,280 females. Two sets contributed only limited numbers of subjects. Therefore, data were pooled and residuals derived for each centre, the differences again being within the range compatible with sampling variability. The maximum deviation in z-scores for FEV₁ and FVC was -0.22 in males (93 individuals in dataset) and -0.68 in females (12 individuals in dataset). Differences in FEV₁ and FVC from the overall mean were proportional (maximum deviation of z-score for FEV₁/FVC between centres was -0.04).

North Africa and Iran

Six datasets [30–37] were available comprising 7,273 subjects (54.1% male) aged 5–90 yrs. There were no significant differences between centres in predicted mean for FEV₁ and FVC, but FEV₁/FVC in Iranian data (n=6,137) was systematically higher than in any other dataset, so they could not be fitted into any group. For example, predicted FEV₁/FVC (5th centile) in North African and Iranian females aged 60 yrs and 168 cm were 0.79 (0.67) and 0.82 (0.74), respectively; in 60-yr-old males (180 cm) the corresponding values were 0.78 (0.66) and 0.82 (0.74), respectively.

TABLE 1 Summary of datasets included in the initial analysis[#]

Group	Countries	Males		Females	
		N	Age range yrs	N	Age range yrs
African-American	1	1529	6–85	2029	6.1–87
India and Pakistan	2	2837	4–86	3003	3–79
Latin America	5	2337	6.7–89.4	2578	7.4–89.7
Mexican-American	1	1622	6.2–86	2282	6.5–87
Iran	1	3398	5–85	2739	5–80
Oman	1	638	6–65	618	6–65
North East Asia	2	2176	15.3–91	4526	15.5–90
South East Asia	4	4187	3.3–88	6371	3.1–92
North Africa	2	541	6–78	602	6–90
Caucasian	14	24229	2.5–95	28844	2.5–95
Other		199	6.2–93	474	5.8–91
Total	33	43693	2.5–95	54066	2.5–95

[#]: total sample included 97,759 subjects.

Indian subcontinent

Data on children were available from India [38–40] and Pakistan (personal communication; G. Mustafa), and on adults of Asian Indian descent from the USA (total $n=5,477$) (table E3 in the online supplementary material) [41]. One of the sets produced a pattern for FEV₁, FVC and FEV₁/FVC that differed significantly from any other dataset, and the remaining two sets of adults and children did not join well. Therefore, reference equations for the Indian subcontinent could not be derived for the current report.

Oman

Data on 1,256 lifelong nonsmokers (51.2% males) aged 6–65 yrs were available [42, 43]. They did not fit into any of the four groups that were formed and could not be combined with data from Iran, therefore, they were not included in the present prediction equations (table E3 in the online supplementary material).

Japan

Although yet to be published, and hence not available to the GLI at time of going to press, a large Japanese dataset (17–95 yrs) has recently been collected as part of the Japan Lung Health Survey (personal communication; M. Kubota, Respiratory Medicine, School of Medicine, Kitasato University, Minato, Japan). Given the “dual-origin hypothesis”, according to which ancestral Japanese populations were brought by two major pre-historic migration events into Japan from the north and south of the Asian continent [44, 45], predicted values may fall between those for North and South East Asians. It is therefore provisionally recommended that predicted values for Japanese subjects are based on the “other” GLI 2012 equation until a suitable coefficient can be developed (see Other section below).

Pooling groups

We set the criteria for combining groups as: 1) minimal effects on predicted mean values; and 2) small effects on the clinically important LLN. The latter was checked by inspection of the percentage of subjects in groups alluded to previously whose observations fell below the 5th centile (LLN 5%, z -score -1.64). As part of this study we found that the smaller the sample size in datasets from people of European ancestry, the more the average and LLN may differ from that of pooled data [14]. However, due to the large size of the overall dataset, exclusion of any small sub-sets ($n < 150$) wherein average z -score deviated by $> \pm 0.4$ from the overall mean, led to trivial ($< 0.01 z$) changes in the average or standard deviation for FEV₁ or FVC. Data were not available from all regions of the world, however; some had to be excluded due to conflicting results. Thus, currently, four groups could be formed: 1) Caucasian: Europe, Israel, Australia, USA, Canada, Mexican–Americans, Brazil, Chile, Mexico, Uruguay, Venezuela, Algeria and Tunisia; 2) Black: African–American; 3) South East Asian: Thailand, Taiwan and China (including Hong Kong) south of the Huaihe River and Qinling Mountains; and 4) North East Asian: Korea and China north of the Huaihe River and Qinling Mountains.

There are practical advantages to combining groups with very similar predicted values, but these must be justified

physiologically and statistically. For example, we can corroborate that predicted values for Mexican–Americans and Caucasian USA citizens are the same [46].

Obviously, there were differences in the mean predicted value and the 5th centile between datasets within a group. As shown earlier such differences decreased as sample size increased [14]. When producing larger subsets from Caucasians, such as Latin Americans, Mexican–Americans and data from North Africa, there was fair agreement in the predicted levels and the 5th centiles, so that these combinations are clinically useful. Several datasets did not fit in any of the groups. Data from Iran [30, 32] had unusually high FEV₁/FVC ratios. Children from Mexico City ($n=4,009$) produced about 8% larger predicted values for FEV₁ and FVC than other sets in the group of Caucasians and, therefore, could not be included. Data on Latin Americans are, therefore, limited to those > 40 yrs old. This left 74,187 subjects in four groups (table 2 and tables E2 and E3 in the online supplementary material).

Other

In the reference equations there are two distinct adjustments for each of the four ethnic groups, one relating to the mean and the other to the CoV (see Prediction Models and Proportional difference between ethnic groups sections). Both are multiplicative adjustments and can be viewed as percentages by multiplying them by 100. From these we have formed an “other” ethnic group, corresponding to groups other than the four main groups, and individuals of mixed ethnic origin. This composite group takes as its mean and CoV adjustments the corresponding adjustments for the four main ethnic groups, averaged over group and sex. Thus individuals in this “other” group are compared to the average of the four main ethnic groups. This should facilitate interpretation until a more appropriate solution is developed.

For ethnic groups not covered by the GLI equations, it will be possible to derive suitable ethnic mean and CoV adjustments without the need to re-calculate the GLI equations. We suggest that a representative sample of at least 300 subjects be used for this, collected using standardised protocols. GLI software provides the facility to incorporate these adjustments for additional datasets as they become available.

TABLE 2 Subjects in the groups for which prediction equations were derived[#]

Group	Males		Females		Total
	N	Age range yrs	N	Age range yrs	
Caucasian	25827	2.5–95	31568	2.5–95	57395
African–American	1520	6–85	2025	6.1–87	3545
North East Asian	1414	16–91	3578	16–88	4992
South East Asian	3095	3.3–86	5160	3.2–92	8255
Total	31856		42331		74187

[#]: the countries of origin are listed in table E1 in the online supplementary data.

Prediction models

The best fitting models required log transformation of height, age, FEV₁, FVC and FEV₁/FVC. The group by age interaction terms yielded insignificant coefficients in most analyses; when coefficients were statistically significant, the effect on predicted values was limited to a few mL. Therefore, the final analyses are based on a model omitting the interactions, and assuming proportional differences between groups (see above):

$$\log(Y) = a + b \times \log(H) + c \times \log(A) + \text{age-spline} + d \times \text{group}$$

where group takes a value of 1 for Caucasians, African-Americans and North or South East Asians, as appropriate (refer to the online supplementary material), and the coefficient d differs between groups. A smoothing age spline was invariably required for predicting the mean and CoV for FEV₁, FVC, FEV₁/FVC, FEF_{25-75%} and FEF_{75%}; it was also required for modelling skewness, except for FEV₁, FVC and FEF_{25-75%} in males and for FVC, FEV₁/FVC and FEF_{25-57%} in females. For FEV_{0.75} and FEV_{0.75}/FVC in the 3–7-yr age range a simple linear model (*i.e.* without look-up tables) sufficed.

Note that the d coefficients for mean and CoV in each ethnic group correspond to the corresponding adjustments (see Other section above).

Simplifying look-up tables

As delineated in the Prediction Model and Entering height and age into the equations sections on the formula for predictive models, a term spline which varies with age arises from fitting a smoothing spline. This is presented as a look-up table for a series of ages which allow interpolation to exact age. It was possible to replace the look-up table for those aged ≥ 25 yrs with an equation, without loss of accuracy, so that lung function devices with limited memory can significantly improve on the use of resources (refer to the online supplementary material). The 3–25-yr age range must still rely on look-up tables. Particularly in preschool children and adolescents such tables need to be quite detailed (to at least 1 decimal age in yrs), as a few months age difference can affect the predicted values by up to 8.5% [47]. Look-up tables are available (www.lungfunction.org/files/lookuptables.xls).

Proportional differences between ethnic groups

The percentage differences between ethnic groups are shown in table 3 and illustrated in figure 2. Except for FEV₁/FVC in South East Asians, predicted values were highest for Caucasians. The FEV₁ and FVC in African-Americans and North East Asians differed from those in Caucasians by the same percentage, signifying that for the same age and height, lung dimensions differed proportionately (table 3).

Effect of collation on predicted values and LLN

Since there were so many more data on Caucasians, including large recently obtained datasets, the analysis of effects of collation on predicted values and their LLN was limited to Caucasians. Predicted values from the five largest recent studies (n=24,783) with acknowledged good quality control [17, 29, 48–50] were compared with results from the collated dataset. The CoV for these five studies was calculated by expressing the residual standard deviation ((predicted - LLN)/1.644) as a per cent of the predicted value. With the exception

of NHANES III and NHANES IV [17, 29], which produced nearly identical values in Caucasians, the predicted values varied between the five large studies, with data from studies within the same country being more different from each other than from the other countries (fig. 3 and fig. E8 in the online supplementary material); this has been attributed to the use of different equipment [50], an important observation highlighting the fact that in spite of good quality control, differences between instruments affect measurement results. Importantly, the CoV for the collated dataset were in between the four larger studies, signifying that collation neither over-inflates variability nor lowers the LLN as a result of poor quality data (fig. 3 and fig. E8 in the online supplementary material). Datasets with unusually high or low average predicted values might influence both predicted mean and scatter. As discussed in the Pooling groups section, removing sets in which the average z-score deviated by $> \pm 0.4$ (*i.e.* $> \sim 5.5\%$ difference) from the overall mean, led to trivial changes in predicted values and their LLN.

Age-related change in pulmonary function

The spirometric indices in this study are a power function of height and age. Hence age-related changes in predicted values are smaller in absolute terms (but not percentage wise) in short rather than tall people. This is shown in figure 4 for FEV₁ in adult Caucasian males of height 160 cm, 175 cm and 190 cm (similar results for FVC). In non-Caucasians predicted values are smaller than in Caucasians of the same standing height; therefore, for the same age and height, the annual cross-sectional change will be smaller than in Caucasians. For example, the annual cross-sectional fall in FEV₁ in an adult African-American is 15% smaller than in a Caucasian male of the same age and height (table 3). As FEV₁ \sim height^k (as is FVC), where k is an allometric constant, the proportionality of age-related cross-sectional changes can be illustrated by standardising the index for height (FEV₁/H^k). Although males and females have different lung volumes, the cross-sectional pattern after standardising for height is very similar (fig. 4). Please note that longitudinal changes differ from cross-sectional age differences [51–63], and are affected by changes in body weight [64–68].

Measured values are often expressed as % predicted, and 80% adopted as the LLN. The scatter around predicted, and hence the LLN, is age-dependent (fig. 5), such that the LLN is not constant but varies depending on the age and the outcome (figs 6 and 7).

Entering height and age into the equations

The prediction equations have the form:

$$\text{predicted value} = e^a \times H^b \times A^c \times e^{d \times \text{group}} \times e^{\text{spline}}$$

where a is the intercept, H is height (cm), b is the exponent for height, A is age (yrs) and c is the exponent for age, and spline is the contribution from the age spline; group is Caucasian, African-American, South or North East Asian, and takes the value of 1 for the appropriate group and 0 for the other groups. It follows that the accuracy of the values entered for height and age matters. For example, the value of k for FEV₁ in males equals 2.22. Entering a H of 160 cm instead of 161 cm changes the predicted value by 1.3%. The calculation of the error for age

TABLE 3 Percentage difference in mean pulmonary function by sex and ethnic group compared to Caucasians

Group	Females				Males			
	FEV ₁	FVC	FEV ₁ /FVC	FEF _{25-75%}	FEV ₁	FVC	FEV ₁ /FVC	FEF _{25-75%}
African-American	-13.8	-14.4	0.6	-11.7	-14.7	-15.5	0.8	-12.9
North East Asian	-0.7	-2.1	1.1	-7.7	-2.7	-3.6	0.9	-3.2
South East Asian	-13.0	-15.7	2.9	-4.3	-9.7	-12.3	2.8	-0.9

FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; FEF_{25-75%}: forced expiratory flow at 25-75% of FVC.

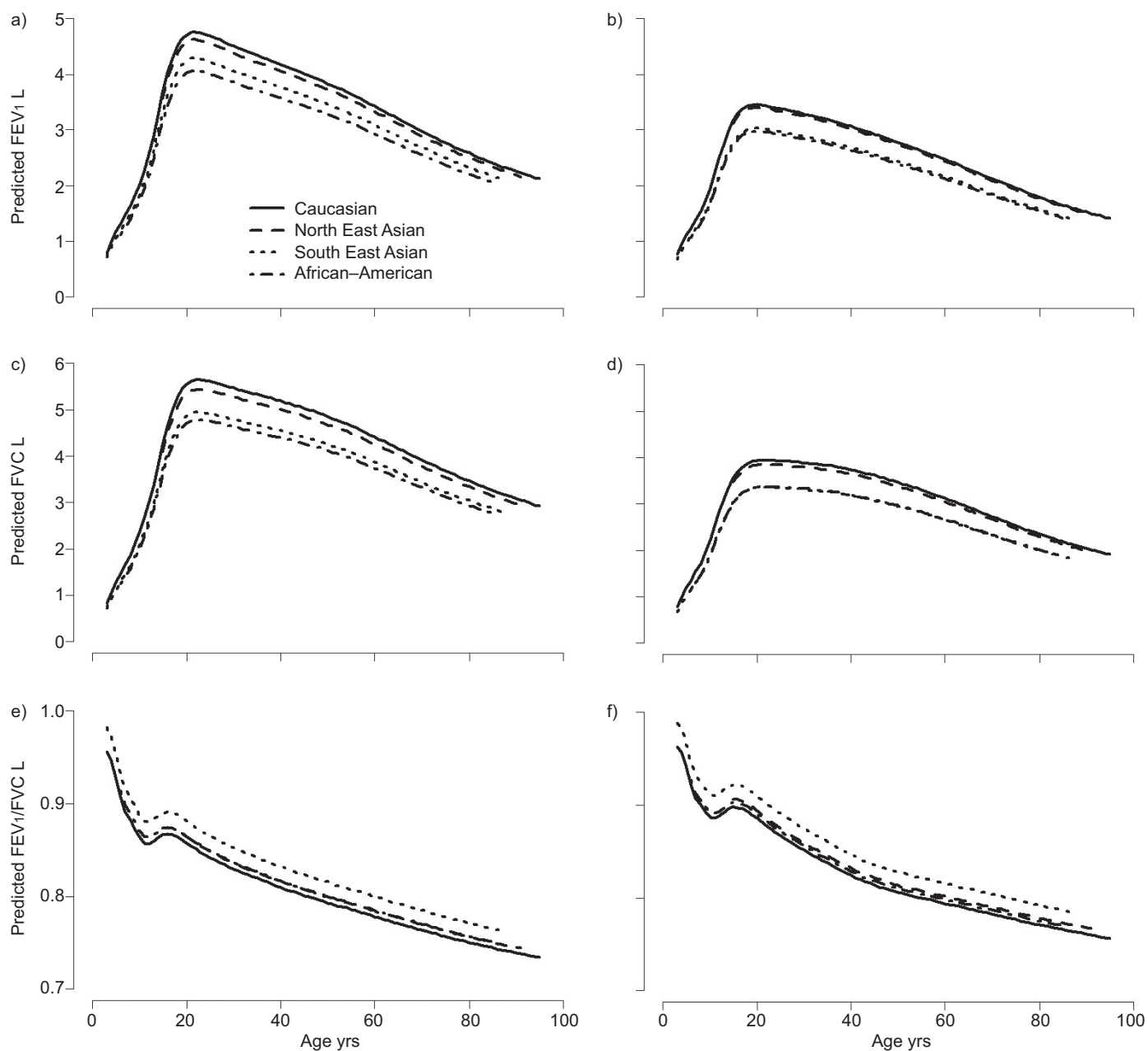


FIGURE 2. Predicted values for a, b) forced expiratory volume in 1 s (FEV₁), c, d) forced vital capacity (FVC) and e, f) FEV₁/FVC by sex and ethnic group. a, c, e) Males and b, d, f) females. Graphs were generated using mean height for age in Caucasians to illustrate proportional differences between ethnic groups of the same height and age; in practice, differences in height for age further affect predicted values. The rise and fall in FEV₁/FVC around adolescence is due to differential changes in FEV₁ and FVC [15].

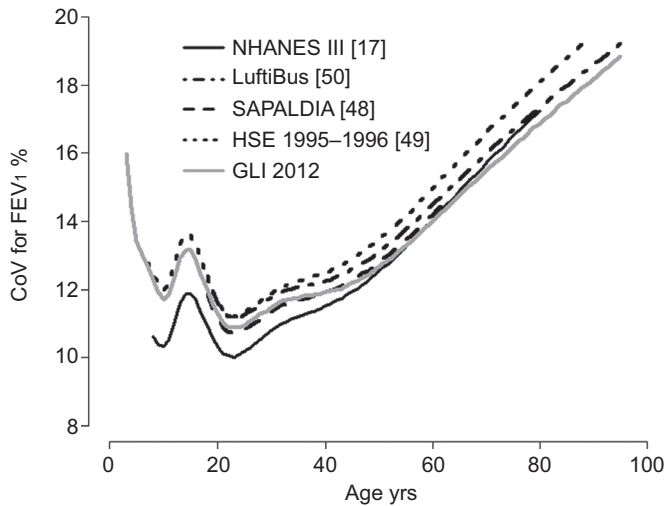


FIGURE 3. Coefficient of variation (CoV) for forced expiratory volume in 1 s (FEV₁) in healthy Caucasian males in four large recent studies, and in the collated Global Lungs Initiative (GLI) material. Graphs were generated using mean height for age in Caucasians to illustrate proportional differences between ethnic groups of the same height and age; in practice, ethnic differences in height for age further affect predicted values. Findings for National Health and Nutritional Examination Survey (NHANES) IV [29] and NHANES III were practically identical. SAPALDIA: Swiss Cohort Study on Air Pollution and Lung Heart Diseases in Adults; HSE: Health Survey for England.

is less straightforward, as the contribution of the spline function varies with age. Errors from entering age inaccurately are largest during adolescence. For example, if one substitutes 14 yrs into the equation instead of 14.9, the predicted volume will be under-estimated by 4.7%. The combined effect of the above 1-cm error in height and 0.9-yr error in age in a 14.9-yr-old young male of height 161 cm will lead to an error in the predicted value of 6%. Self-reported height may differ by as much as 6.9 cm from measured height [69–75]. It follows that one should not rely on self-reported height, and that it should be measured using a calibrated stadiometer; actual age should be entered with one decimal point of accuracy.

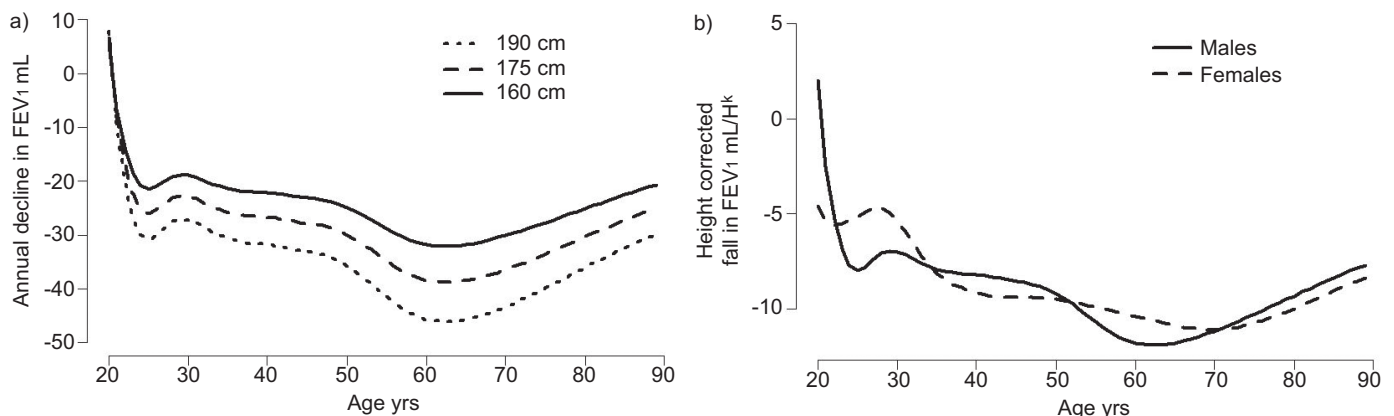


FIGURE 4. a) Age-related cross-sectional annual change in forced expiratory volume in 1 s (FEV₁) is larger in tall subjects rather than short subjects. This shown for Caucasian males of height 160 cm, 175 cm and 190 cm. b) Standardising FEV₁ for height by taking the ratio FEV₁/height^k, where k is ~2, largely removes differences in age-related changes between adult males and females. These findings have no bearing on within-subject longitudinal decline.

The look-up tables for the spline in age have a resolution of 0.25 years, *i.e.* 3 months. In the example above, a 3-month error in actual age may still lead to a 0.9% error in the predicted value. Errors can be virtually eliminated, and the look-up table appreciably reduced in size, by obtaining the value for the smoothing spline in age by interpolating between the two nearest ages in the look-up table. From 25 yrs of age onwards, the look-up table can be replaced by equations which, if used, avoid the need for interpolation and loss of accuracy (refer to online supplementary material).

DISCUSSION

This is the first study to present spirometry prediction equations spanning ages 3–95 yrs for ethnic and geographic groups from 26 countries. It is the result of unprecedented, unselfish and professional international cooperation endorsed by six international societies. As the results show, however, it is only the first leg of a journey.

An important benefit of the availability of many data from so many sources is that results are more generalisable across populations. As shown in this study, subsets of children and adolescents generate predicted values that connect smoothly to those from subsets comprising only of adults. This implies that our prediction equations cover the entire age range, even in countries that contributed data covering a limited age range. Another advantage is that patterns that emerge from one study can be validated in others. For example, a previously unrecognised pattern in FEV₁/FVC was identified in children and adolescents, where the ratio increased during adolescence rather than decreasing monotonically from childhood to adulthood. This is because in childhood, FVC outgrows the total lung capacity and FEV₁, leading to falls in FEV₁/FVC, a trend that is reversed in adolescence, a pattern that went undetected with conventional statistical techniques (fig. 2) [15]. By undertaking these analyses, it was possible to confirm that this is a physiological pattern of development reflecting differential growth of FEV₁ and FVC during adolescence, common to all ethnic groups. Similarly, whereas it might have been speculated that, within any specific ethnic group, differences in the level and variability of measurements from participating centres arose from differences in standards or secular trends (*i.e.* a trend in

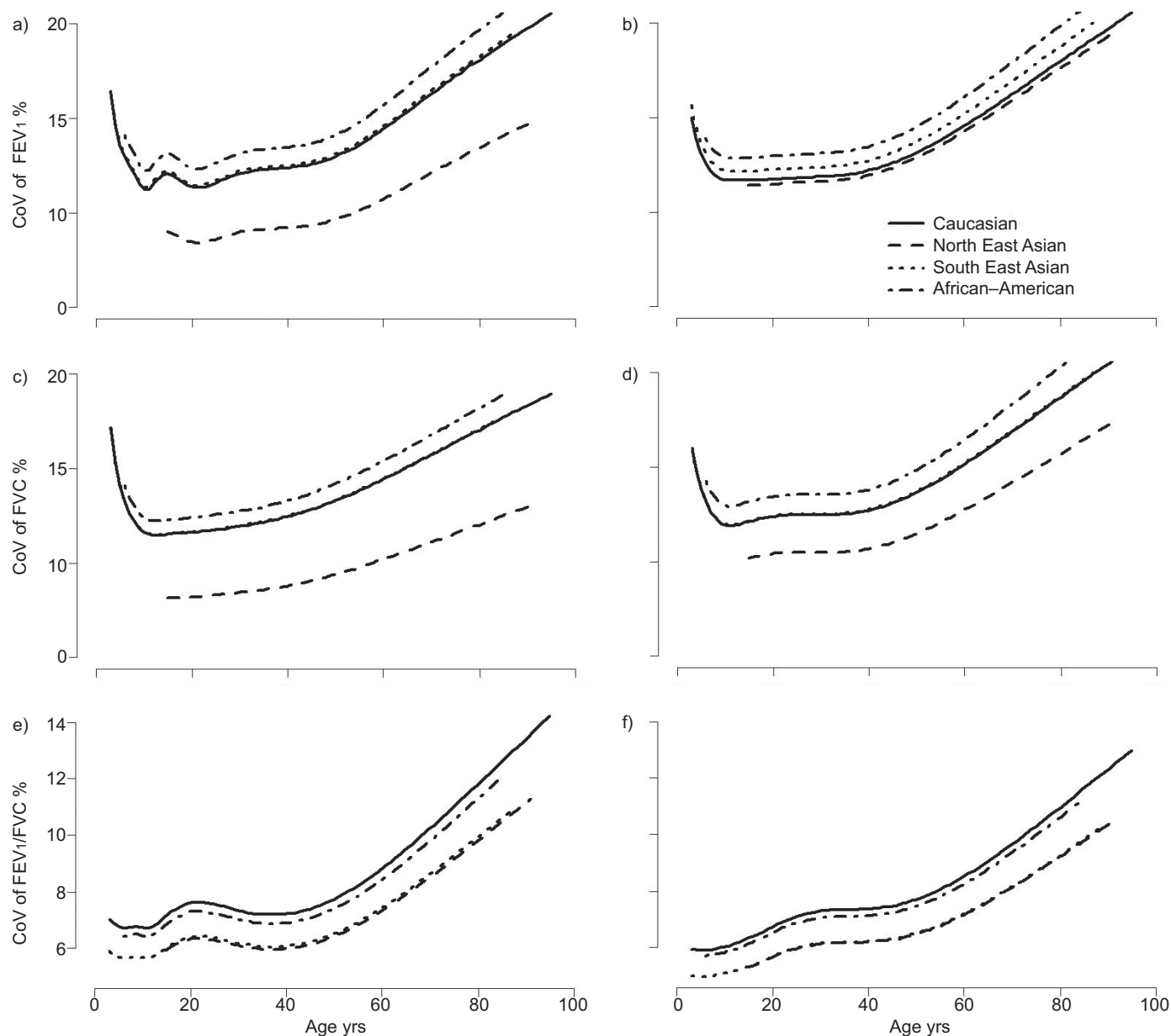


FIGURE 5. Between-subject coefficients of variation (CoV) of a, b) forced expiratory volume in 1 s (FEV₁), c, d) forced vital capacity (FVC) and e, f) FEV₁/FVC in Caucasians, African-Americans, and North and South East Asians by age. a, c, e) Males, and b, d, f) females.

lung function with the year of birth), clear evidence has been provided that this did not affect our findings [14].

This study confirms the existence of proportional differences in pulmonary function between ethnic groups (table 3) [76], signifying proportionate scaling of lung size due to differences in body build, so that the FEV₁/FVC ratio is generally independent of ethnic group. This has clinical advantages in that, with the exception of South East Asians, in whom the FEV₁/FVC ratio is 2.6–2.8% higher than in other groups, it allows a uniform definition of airways obstruction (*i.e.* pathological airflow limitation) based on the LLN for FEV₁/FVC across ethnic groups. Whereas the FEV₁/FVC ratio in Caucasians can potentially be applied to any group of subjects with reasonable

confidence, this does not apply to the other spirometric outcomes. When assessing lung function in an individual not represented in one of the present four groups, the equation for the “other” group is available, although the output generated by pulmonary function equipment must then alert the user that the predicted values (for FEV₁, FVC, *etc.*) may not be appropriate for the subject, and that the results should be interpreted with caution. This may be particularly true for individuals from the Indian subcontinent, for whom published literature suggests that predicted values may be at least as low as those found in Black subjects. With ongoing data collection it is hoped that the majority of “missing groups” will be included within the next 5 yrs. As discussed in the Other section, provided a sufficiently large and representative dataset is collected using standardised protocols, it

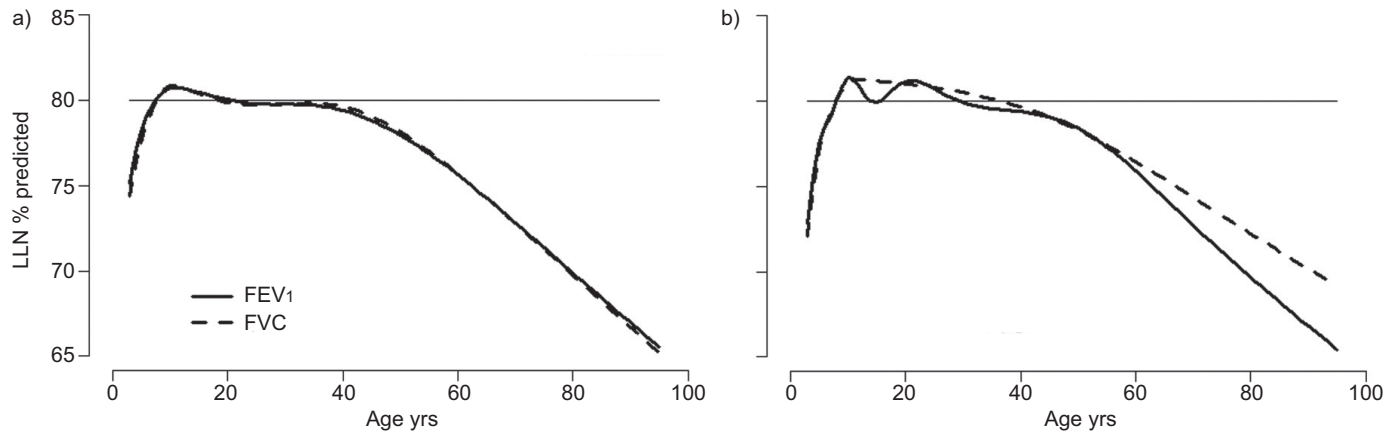


FIGURE 6. The lower limit of normal (LLN), based on the 5th centile (-1.64 z-scores), values for forced expiratory volume in 1 s (FEV₁) and forced vital capacity (FVC) in a) Caucasian females and b) males. As can be seen, the use of 80% of predicted as the LLN gives rise to age-related bias. Graphs were generated using mean height for age in Caucasians.

should be possible to derive suitable ethnic coefficients for specific groups without re-calculating the GLI equations.

At present, it is common practice to standardise measurements for differences in height, age and sex by converting them to % predicted value; however, this approach has important clinical consequences, particularly when used to classify patients and disease severity using a fixed cut-off [77–81]. STANOJEVIC *et al.* [12] were the first to show that the coefficient of variation for FEV₁ and FVC varies with age, the greatest variability occurring in young children and the elderly, with a minimum during early adulthood. This study confirms and extends this observation by showing that it occurs in all ethnic and geographically defined groups (fig. 5), including a temporary increase in variability during the adolescent growth spurt. This pattern was observed in individual studies with a large sample size and with known

good quality control (fig. 3), further supporting that this is not an artefact.

Representativeness of equations for various groups

Comparison of predicted values in four large recent studies [17, 48–50] with the present study shows that there is no inflation of the coefficient of variation, and that, if anything, the predicted values are marginally higher (table 4 and fig. E8 in the online supplementary material). These findings confirm that the new equations represent measurements obtained with good quality control, encompassing the entire process from choosing a representative population sample of healthy life-long nonsmokers down to selecting laboratory procedures.

Pooling data on the basis of ethnicity proved to be a valid starting point. Taking into account chance differences due to varying sample size, datasets within groups agreed well, so

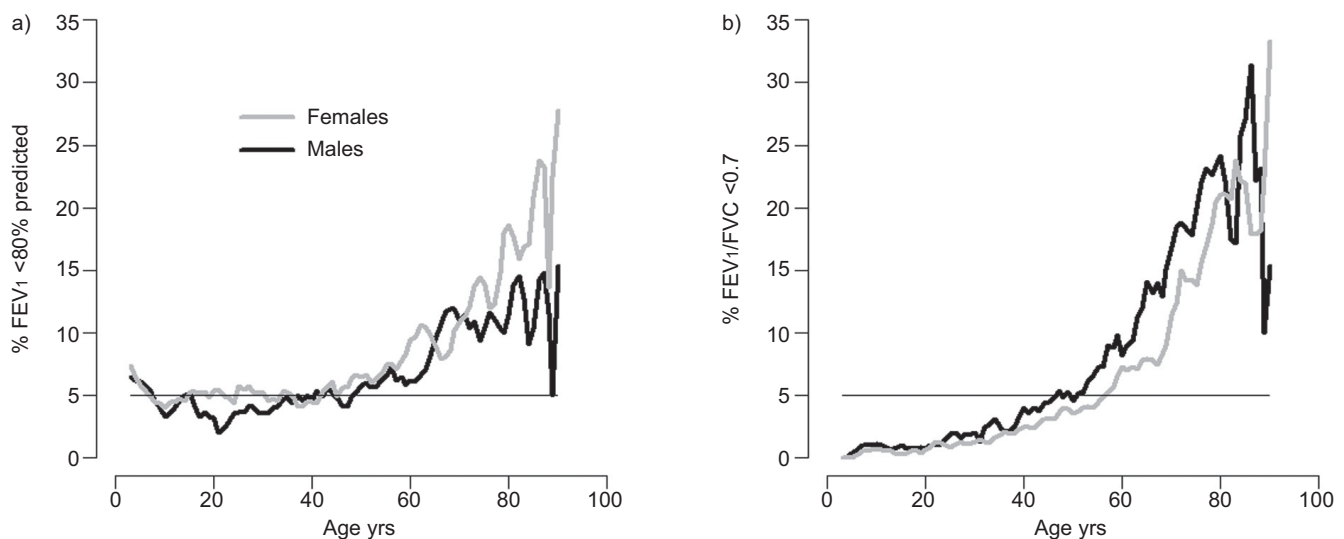


FIGURE 7. Percentage of healthy male and female nonsmokers ($n=74,187$) in whom a) forced expiratory volume in 1 s (FEV₁) is <80% predicted or b) FEV₁/forced vital capacity (FVC) is <0.70. If it is assumed that FEV₁/FVC is <0.70 (instead of below the 5th centile lower limit of normal) implies pathological airflow limitation, cases are missed in those <50 yrs of age, whereas over diagnosis occurs above that age. Thus, at 80 yrs of age, 20–25% of the reference population would have low values compatible with pathological airflow limitation instead of the expected 5%; this represents 75–80% false-positive test results (b).

TABLE 4 Comparison of predicted values in adult Caucasian males according to different authors

Author [ref.]	a) 17.9 yrs, 160 cm			Author [ref.]	b) 18.0 yrs, 160 cm		
	FEV ₁	FVC	FEV ₁ /FVC		FEV ₁	FVC	FEV ₁ /FVC
POLGAR [7]	3.09	3.42	NA	ECSC/ERS [8]	3.67	4.23	0.83
ROSENTHAL [82]	2.70	3.69	0.84	HANKINSON [17] [†]	3.58	4.12	0.84
ZAPLETAL [83]	2.93	3.53	0.85	STANOJEVIC [13] [‡]	3.46	4.09	0.85
STANOJEVIC [13] [‡]	3.45	4.08	0.85	GLI 2012	3.61	4.13	0.88
WANG [6]	3.49	3.93	0.87				
GLI 2012	3.61	4.12	0.88				
Author [ref.]	c) 17.9 yrs, 180 cm			Author [ref.]	d) 18.0 yrs, 180 cm		
	FEV ₁	FVC	FEV ₁ /FVC		FEV ₁	FVC	FEV ₁ /FVC
POLGAR [7]	4.30	4.69	NA	ECSC/ERS [8]	4.53	5.38	0.83
ROSENTHAL [82]	4.27	5.17	0.82	HANKINSON [17] [†]	4.53	5.39	0.84
ZAPLETAL [83]	4.11	4.99	0.85	STANOJEVIC [13] [‡]	4.56	5.45	0.85
STANOJEVIC [13] [‡]	4.55	5.44	0.85	GLI 2012	4.69	5.49	0.86
WANG [6]	4.46	5.20	0.86				
GLI 2012	4.68	5.47	0.86				
Author [ref.]	e) 25 yrs, 175 cm			Author [ref.]	f) 55 yrs, 175 cm		
	FEV ₁	FVC	FEV ₁ /FVC		FEV ₁	FVC	FEV ₁ /FVC
CRAPO [84]	4.45	5.32	0.84	CRAPO [84]	3.71	4.67	0.79
ECSC/ERS [8]	4.31	5.09	0.83	ECSC/ERS [8]	3.44	4.31	0.77
HANKINSON [17] [†]	4.44	5.36	0.83	HANKINSON [17] [†]	3.63	4.74	0.77
HSE [49]	4.43	5.29	0.85	HSE [49]	3.60	4.63	0.79
KNUDSON [85]	4.39	5.24	0.84	KNUDSON [85]	3.52	4.35	0.81
STANOJEVIC [13] [‡]	4.42	5.36	0.83	STANOJEVIC [13] [‡]	3.63	4.75	0.77
GLI 2012	4.46	5.32	0.85	GLI 2012	3.65	4.66	0.79

FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; GLI: Global Lungs Initiative; ECSC: European Community for Steel and Coal; ERS: European Respiratory Society; HSE: Health Survey for England 1995–1996; NA: not available. [‡]: all age; [†]: National Health and Nutritional Examination Survey (NHANES) III. The transition from one set of equations to another on the 18th birthday (a versus b or c versus d) is not smooth using older equations. The discontinuity, and the disparity between predicted values is much worse in those who are short for their age (a versus b), than in those of average height (c versus d). In adult males there is fair agreement between predicted values, although those derived from the ECSC/ERS are systematically lower than others for subjects ≥ 25 yrs of age (e versus f).

that the present set of equations forms a firm basis for application in research and clinical practice. Yet, there are limitations. The differences observed between the Indian datasets may arise from considerable variation in socio-economic conditions, particularly in children and young adults [40, 86, 87]. The 1 billion inhabitants of the Indian subcontinent comprise a large number of ethnic groups from various racial strains [88] from South to North; socioeconomically and ethnically the population is heterogeneous. Given the large differences observed, the Task Force cannot currently make a recommendation for predicted values for this part of the world, although the decrement may be at least as great as that observed in Black subjects. More data, including information about ethnicity, is required. Although the data from Latin American adults appeared to fit in with the other Caucasians, this was not the case for children and adolescents in Mexico City (Mexico), who produced higher predicted values for FVC and FEV₁. This has been reported previously [21]. It could not

be attributed to being born and raised at altitude (2,250 m), but has been attributed to the fact that they have shorter legs for stature compared to other ethnic groups [89, 90]. Latin Americans are a mix of people from Spanish descent and a spectrum of indigenous people, often living at high altitudes. There is also considerable intermarriage. The predicted values should, therefore, not be applied indiscriminately to Latin Americans of non-European descent.

This study shows that there are proportional differences in the level of pulmonary function between Caucasians, South and North East Asians, and African-Americans (table 3). These differences are compatible with the “out of Africa” theory. This theory, supported by information from genetic markers [91], points to mankind originating in Africa and migrating through South East Asia, travelling northwards and slowly populating East Asia while undergoing evolutionary changes. Our findings also confirm a previous report about differences

in pulmonary function between North and South China [92–94]. When comparing findings in adults in Hong Kong to those of HANKINSON *et al.* [17] (NHANES III), differences in FVC varied between 12 and 17% depending on age and sex [23], similarly to the present findings (table 3). However, CRAPO *et al.* [95] found average FVC and FEV₁ in native Mongolians to be within 1–2% of the Caucasian predicted values, as in the present study of North East Asians (table 3). In view of differences in socioeconomic conditions between urban and rural communities in China [93, 94, 96–99], differences between Han people and minority groups [91, 93, 96, 97] and ethnic differences in body build [99–101], the present prediction equations may not fit all East Asian ethnic groups equally well. Whereas post-war improvements in socioeconomic conditions in Japan led to increased height driven by growth of leg length [102], with the exception of Chinese girls in Beijing [103], there is no obvious secular trend in the leg length/height ratio [100, 103, 104], suggesting that as people grow taller lung volumes increase proportionately. However, more information is required about pulmonary function in other ethnic groups, including ethnic minorities.

Data quality

A major concern when collating data is that the selection of subjects, measurement techniques including measurement of stature, differences between instruments, laboratory standards and quality criteria when selecting spirometric indices may lead to predicted values and LLN which are biased due to inclusion of poor quality data. All the above factors may, to a varying extent, have influenced the present results. Rigorous *post hoc* quality control of each aspect might have helped to minimise such contributions to variability, but this was not feasible for >150,000 sets of results. However, with few exceptions the data had previously been published in peer-reviewed journals, and complied with temporal international recommendations. Evidence for unsatisfactory quality of individual datasets might come from outlying values for predicted values, and from variability. When testing this in Caucasians, large datasets from different parts of the world agreed remarkably well; the smaller the dataset, the larger the spread of deviations from the overall mean [14]. This reflects the fact that the smaller the sample, the greater the likelihood that it is not entirely representative of the population. Thus, there is normal variability between population samples and no evidence that predicted values and the LLN are biased by poor quality studies. By implication, small samples are not appropriate for validating reference equations for use in individual laboratories; a minimum of 150 males and 150 females is required [14], making this effort impracticable for most laboratories. Further evidence that poor data quality has not affected the present study is that predicted values and LLN are, for practical purposes, the same as in four large recent studies with good quality control (fig. 3 and fig. E8 in the online supplementary material) [17, 48–50].

Quality control of spirometric data, applying temporal international standards, is a double-edged sword. In a study of adults carried out in a laboratory [105] that adhered to the quality assurance programme now widely adopted in New Zealand and Australia [106], >30% of measurements needed to be discarded because they did not meet ATS quality criteria [107]. The vast majority of these exclusions related to subjects

aged <30 yrs; the major stumbling block was the requirement that subjects should exhale for at least 6 s. It illustrates that present day quality criteria are not compatible with what is achievable in professionally run laboratories, and need reviewing so as to be uniformly applicable in the field. Strict adherence to the present end of test criteria will lead to biased datasets, for example by favouring young adults with the longest forced expiration times and hence, in all likelihood, a smaller FEV₁ and FVC than their counterparts who consistently empty their lungs within 6 s. Similar reasoning applies to children. It is unclear to what extent such bias crept into this study. Because an FVC can always be obtained, unlike FEV₆, the present prediction equations are applicable across all ages. For this reason, FEV₁/FVC may be a more appropriate outcome than FEV₁/FEV₆ when diagnosing obstructive lung disease across a wide age range.

Secular trends

In Caucasian subjects, with data collected over three decades, no evidence was found for a secular trend (*i.e.* a trend in the level of lung function in successive birth cohorts) in pulmonary function [14]. However, such trends may well emerge as more data from developing countries become available, where improving socioeconomic conditions lead to greater physical development and improved health.

Age-related changes in FEV₁ and FVC

As FEV₁ and FVC are power functions of height and age, the cross-sectional annual change in these variables is larger in tall rather than short people (fig. 4). For example, in a Caucasian adult male there is an accelerating cross-sectional decline in FEV₁ after the age of 30 yrs, with a nadir at age 62 yrs when the annual loss ranges between 32 and 46 mL. The decline then decelerates, possibly reflecting a healthy survivor effect.

Mixed ethnic descent

The well-documented ethnic and racial differences in pulmonary function arise from differences in body build (such as chest size or the ratio of sitting to standing height), socioeconomic status (which determines bodily development in early life and leads to secular trends in body size and pulmonary function), growing up at altitude, and possibly other environmental factors [108–119]. In the present study, race and ethnicity were self-reported, which may not be accurate enough for clinical purposes [120–122]. Indeed, in the absence of genetic typing, predicted values in self-reported African-Americans may be biased by up to 200 mL [123]. This may have bearing on clinical diagnosis in view of the association between genetic make-up and disease [124–130]. Until genetic typing becomes standard practice and prediction equations incorporating genetic information become available, lung size for mixed ethnicity can only be based on the ethnicity of the parents, or by using the “other” group and interpreting results in the context of clinical symptoms.

Application in clinical practice

For any given height and sex, a 1-yr difference in age can alter the predicted value by up to 8.5% [47] in those aged <20 yrs. Many commercial devices round age down to completed years, a practice which is unacceptable as it introduces significant bias, especially in young children. To minimise errors age should be recorded with decimal accuracy (preferably as date of measurement minus date of birth). In paediatrics, many frequently used

prediction equations ignore age completely [6, 7, 82, 83], which is also likely to lead to bias.

The FEV₁/FVC ratio in young children can be very high, with the predicted value, and particularly the ULN, exceeding 1.0. In such cases they should be truncated at 1.0. Predicted values for African-Americans and East Asians differ proportionally from those in Caucasians in two respects: 1) the absolute values for FEV₁ and FVC differ by a certain percentage; and 2) this percentage is very similar for FEV₁ and FVC. As a result, the FEV₁/FVC ratios in these groups were virtually identical (table 3); South East Asians, with the highest FEV₁/FVC ratio, are the exception to this rule (fig. 2).

Caution is required when interpreting the prediction equations for those over aged ~75–80 yrs, where the sample size is small (fig. 1 and table E5 in the online supplementary material). This also holds for North East Asians aged <15 yrs (fig. 1). In addition, the equations should not be applied to indigenous Latin Americans. We found no evidence for differences in pulmonary function among Caucasians in different parts of the world, or between East Asians in the USA and China. However, Indian Asians born and raised in the USA were found to have higher pulmonary function for age and height than those born and raised in India [131]. Hence more studies are required to elucidate the influence of country of birth on the level of lung function.

Clinical decision making

Making a clinical diagnosis is an art, where test results help to confirm or reject the diagnosis. A test result is regarded as compatible with disease if it is outside the normal range, usually defined by the mean \pm 2SD (strictly 1.96), which extends from the 2.5th to the 97.5th centile of the distribution. The ATS and the ERS [8, 80, 132] both recommend the use of the 5th centile to define the LLN (*i.e.* -1.64 z-scores). z-scores indicate how many standard deviations a measurement is from its predicted value. Unlike % predicted, they are free from bias due to age, height, sex and ethnic group, and are therefore particularly useful in defining the LLN and ULN; they also simplify uniform interpretation of test results, particularly if presented as pictograms (fig. 8).

When interpreting multiple tests which are physiologically related, applying the 5th centile LLN to each of them and accumulating the results leads to a high percentage of false-positives. Thus, in 41,136 reference females in this study, the cumulative percentage of test results for FEV₁, FVC and FEV₁/FVC below the 5th centile is 10.4%. Using the 2.5th centile (z-score -1.96) reduces this to 5.6%. In 30,895 reference males the corresponding figures are 10.6 and 5.8%. In view of this, the 2.5th centile LLN (z-score -1.96) is recommended as the decision limit for screening and case finding purposes. However, in subjects with prior evidence of lung disease a borderline low value of FEV₁/FVC, FEV₁ or FVC is more likely to be associated with disease; depending on the strength of clinical evidence of disease, and the cost and consequences of a false-positive or false-negative test result, an LLN at the 5th centile (z-score -1.64) is clinically acceptable.

It is common practice to regard 80% pred as the LLN. However, the true LLN, expressed as % predicted, varies considerably with age (figs 6 and 7). Hence, the use of a fixed

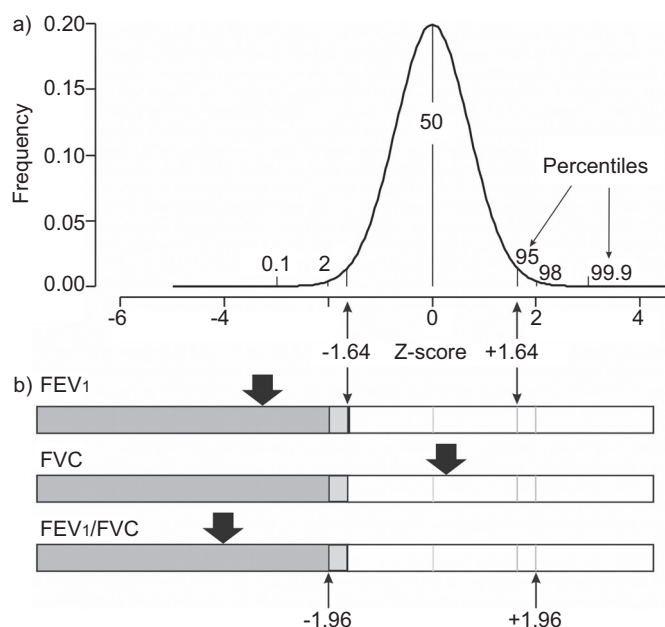


FIGURE 8. a) Illustration of how z-scores relate to percentiles. b) Pictogram showing test results as z-scores (arrows) for a male from North China, aged 34 yrs, height 166 cm. White area in the bars delineates observations that are above the mean -1.64 z-score (lower limit of normal (LLN) 5%) and light grey delineates mean -1.96 z-score (LLN 2.5%). Reproduced from [133] with permission from the publisher.

threshold such as 80% pred for FEV₁ and FVC or 0.7 for the FEV₁/FVC ratio across all ages is discouraged as it leads to significant misclassification (fig. 7) [81]. In the collated data a z-score for FEV₁/FVC below the 5th centile was associated with an FEV₁ below the 5th centile in 1.2% of cases, with FEV₁ % pred <80% in 1.3% of cases, and FEV₁ % pred <70% in 0.4% of cases. In those aged >40 yrs these percentages were 1.2%, 1.6% and 0.7%, respectively. This illustrates that the z-score is not biased by age.

FEF_{25–75%} and FEF_{75%} are measured conditionally on the FVC. Therefore, the between-subject variability in these indices is the sum of intrinsic variability and variability in FVC. Depending on age, sex and ethnic group, the between subject CoV varies between 20 and 62% for FEF_{25–75%} and between 27 and 89% for FEF_{75%} (fig. E9 in the online supplementary material). While the FEF_{75%} and FEF_{25–75%} are not among the indices recommended by ECSC/ERS [8], ATS [107, 132] or ATS/ERS [16], they were included in the current analyses in response to requests from colleagues, especially those caring for children. The very high CoV severely limit the use of these indices for diagnostic purposes in adults, but this does not preclude smaller coefficients of variation within subjects, or the use of these indices in aetiological studies where differences between groups may provide valuable clues. FEV₃ and FEV₆ were not considered as many children and adults cannot exhale for 3 and 6 s, respectively, and because very few data were submitted. In order to improve the current equations for these outcomes, future studies in healthy subjects should ensure these are measured; in view of the large coefficients of variation in those aged <20 yrs (20–33% for FEF_{25–75%} and 27–42% for FEF_{75%}). Further research is also required to determine whether such outcomes are in fact clinically useful.

Strengths and weaknesses

The strengths of this study are that it is based on a large population sample of high-quality data from five continents, analysed using a technique that allows modelling both continuous predicted values from pre-school to old age and their LLN. It is the first time that such equations have been modelled simultaneously for several ethnic groups, confirming the presence of proportional differences between them (fig. 2, table 3) [76, 134]. Results in Caucasians are comparable to those in large recent studies (fig. E8 in the online supplementary material). Data from Africa, Polynesians, the Indian subcontinent, the Arab world and South America are urgently required. Whereas the data from South America seem to fit those for Caucasians, data from children in Mexico City tend to be higher; thus, caution is required in applying them to Latin Americans over the entire age range, or to those living at altitude or of indigenous descent, as they were not included in the study. Finally, more data from the elderly are required in all groups, and from young non-Caucasian children.

This study showed good agreement between predicted values in large studies with good quality control and subject selection. Therefore, the perceived need for locally defined reference limits probably arises more from the lack of standardisation of selection procedures and quality control than from local differences in reference populations. Even if all reasonable precautions have been taken, large samples from the same population do not necessarily yield the same predicted values [48, 50]; collation of good quality studies has the advantage of yielding a very large sample that is more representative of the population [14].

The current GLI 2012 equations, supported by the major international respiratory societies, have the potential to improve the interpretation of spirometry results, and to standardise interpretation across centres and countries. Many manufacturers have already implemented the "all-age" reference equations of STANOJEVIC *et al.* [13] and are, therefore, in an excellent position to update to the present equations, thus avoiding the use of poorly connecting age-grouped prediction equations. For example, frequently used regression equations lead to major disagreements in predicted values in adolescents, and they connect poorly to those in adults (table 4). The transition is particularly poor in children who are short for age, for example due to chronic lung disease. In addition, the use of different equations and moving from one set to the next confounds comparisons of patients from different centres around the world (table 4). Algorithms and stand-alone software for the GLI 2012 equations are freely available from www.lungfunction.org

Continued use of popular older equations is likely to lead to misclassification of patients. This has direct bearing on the classification of subjects with lung disease, allocation of subjects to treatment regimens in intervention studies, on BODE and other multidimensional indices, and studies into the prevalence or natural course of respiratory disease. However, most commercial database software will allow recalculation of predicted and derived values of previously recorded data simply by changing the preferred reference equation, thereby allowing accurate trend reports to be maintained within individual patients and appropriate comparisons within longitudinal epidemiological studies.

Future developments

Given the pivotal role of proper predicted values for pulmonary function, and the recognised limitations of the present study, an update will be required in a few years when more data of non-Caucasian origin will be available. The present and future data form a treasure trove for research. It is therefore recommended to establish a body, preferably under the auspices of international societies, with a long-term commitment to manage new data meeting minimum requirements, and issue periodic updates to the equations. This organisation requires appropriate management of data, including protection of privacy and owner's interests/rights. The organisation should provide an opportunity for the use of data under strict conditions.

CONCLUSIONS

This study has led to the derivation of continuous equations for predicted values and age-appropriate LLN for spirometric indices from 3–95 yrs of age, based on 74,187 records from healthy nonsmoking males and females from 26 countries across five continents. Ethnic and geographic groups can be grouped under the headings Caucasian, African-American, and North and South East Asian. However, information on pulmonary function is still incomplete in many parts of the world, and more data on subjects aged >75 yrs are required. In the meantime, the predicted values and reference range for those over aged ~75–80 yrs in all groups, and in North East Asians aged <15 yrs, should be interpreted with caution. Comparison with other large studies confirms that the collated data are of good quality, with differences being due to sampling error, and that collation of data has not led to inflation of the coefficient of variation. The GLI 2012 reference equations are a huge step forward, providing a robust reference standard to streamline the interpretation of spirometry results within and between populations worldwide.

RECOMMENDATIONS

Widespread use of these global all-age equations depends on timely implementation by manufacturers of spirometric devices, and should be encouraged by users and the international respiratory societies.

Given the extensive data now available from Caucasian subjects aged 3–75 yrs, and the stability of the GLI 2012 spirometry reference equations for such individuals, collection of further Caucasian normative data is not required. Further data are, however, required for those >75 yrs of age, and for preschool children.

More studies are required in non-Caucasians, particularly Arab, Indian, Polynesian, African and Latin American subjects, including ethnic minorities. Such studies should adhere to international guidelines with respect to methodology, quality control and selection of a representative sample of healthy reference subjects aged 3–95 yrs.

For ethnic groups not covered by the GLI equations, representative samples (of at least 300 subjects) can be used to validate use of one of the four groups, and/or create an appropriate coefficient (adjustment factor) for a new group. Provided a sufficiently large and representative sample, collected using standardised protocols, is available, it should

be possible to derive suitable ethnic coefficients without recalculating the GLI equations.

For individuals not currently represented by the GLI 2012 equations, or of mixed ethnic origin, use of the “other” equation is recommended, with the caveat that results must be interpreted cautiously until a more appropriate solution is developed.

An international repository should be established to manage both existing and future datasets to facilitate data sharing and update of the equations when required.

When collecting lung function data or evaluating test results, particularly during periods of rapid growth, age should be recorded in years with accuracy to at least one decimal place (preferably as date of measurement minus date of birth), as should standing height in cm, measured using a calibrated stadiometer.

Defining the LLN as a fixed per cent of predicted FEV₁ or a fixed FEV₁/FVC ratio leads to age-, height-, sex- and ethnic group-related bias, and such definitions should not be used.

The LLN used should be appropriate for the purpose. Hence, if there is prior evidence of lung disease, a borderline value of FEV₁/FVC, FEV₁ or FVC is more likely to be associated with disease and an LLN at the 5th centile (LLN 5%, z-score -1.64) is clinically acceptable.

By contrast, in epidemiological studies and case finding of asymptomatic subjects, where the cost and consequences of false-positive and false-negative test results are over-riding, an LLN corresponding to the 2.5th centile (LLN 2.5%, z-score -1.96) is recommended as the decision limit.

SUPPORT STATEMENT

This report has been endorsed by the European Respiratory Society (ERS), American Thoracic Society (ATS), Australian and New Zealand Society of Respiratory Science (ANZSRS), Asian Pacific Society for Respiratory (APSR), Thoracic Society of Australia and New Zealand (TSANZ) and the American College of Chest Physicians (ACCP). This study includes data from the MESA study, Funded by National Institutes of Health R01-HL077612, N01-HC095159-165, N01-HC095169.

STATEMENT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

The authors' affiliations are as follows. P.H. Quanjer: Dept of Pulmonary Diseases and Dept of Paediatrics, Erasmus Medical Centre, Erasmus University, Rotterdam, the Netherlands. S. Stanojevic: Portex Respiratory Unit, UCL Institute of Child Health, London, UK and Child Health Evaluative Sciences and Respiratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada. T.J. Cole: MRC Centre of Epidemiology for Child Health, UCL Institute of Child Health, London, UK. X. Baur: Universitätsklinikum Hamburg-Eppendorf, Zentralinstitut für Arbeitsmedizin und Maritime Medizin, Hamburg, Germany. G.L. Hall: Respiratory Medicine, Princess Margaret Hospital for Children, and School of Paediatric and Child Health and Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Perth, Australia. B.H. Culver: Division of Pulmonary and Critical Care Medicine, Dept of Medicine, University of Washington, Seattle, WA, USA. P.L. Enright: Division of Public Health Sciences, University of Arizona, Tucson, AZ, USA. J.L.

Hankinson: Hankinson Consulting, Athens, GA, USA. M.S.M. Ip: Dept of Medicine, The University of Hong Kong, Queen Mary Hospital, Hong Kong, China. J. Zheng: Guangzhou Institute of Respiratory Disease, State Key Laboratory of Respiratory Disease, Guangzhou, China. J. Stocks: Portex Respiratory Unit, UCL Institute of Child Health, London, UK.

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 (Institute of Social and Preventive Medicine, Swiss Tropical Institute, Basel, Switzerland) is also gratefully acknowledged. This study is based on data from 70 centres, including the National Health and Nutrition Examination Survey (NHANES) III, NHANES IV and the Multi-Ethnic Study of Atherosclerosis (MESA) studies. The MESA and MESA lung studies are conducted and supported by the National Heart, Lung and Blood Institute in collaboration with the MESA and MESA lung investigators. In addition to review by representatives of all bodies contributing data to the Global Lungs Initiative, this manuscript has been reviewed by the MESA investigators for scientific content and consistency of data interpretation with previous MESA publications and significant comments were incorporated prior to submission. A full list of participating MESA investigators and institutions can be found at www.mesa-nhlbi.org

The members of the ERS Global Lung Function Initiative (www.lungfunction.org) are as follows. Chairs: J. Stocks, X. Baur, G.L. Hall and B. Culver. Analytical team: P.H. Quanjer, S. Stanojevic, T.J. Cole and J. Stocks. Additional members of the Steering Committee: J.L. Hankinson, P.L. Enright, J. Zheng and M.S.M. Ip. Statistical reviewer: C. Schindler.

The following people and centres contributed data to this manuscript. O.A. Al-Rawas (Dept of Medicine, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Sultanate of Oman); H.G.M. Arets (Dept of Paediatric Pulmonology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, the Netherlands); C. Bárbara (Portuguese Society of Pneumology, Lisbon, Portugal); R.G. Barr and the MESA study (Columbia University Medical Center, New York, NY, USA); E. Bateman (University of Cape Town Lung Institute, Cape Town, South Africa); C.S. Beardmore (Dept of Infection, Immunity and Inflammation (Child Health), University of Leicester, Leicester, UK); H. Ben Saad (Laboratory of Physiology, Faculty of Medicine, University of Sousse, Sousse, Tunisia); B. Brunekreef (Institute for Risk Assessment Sciences, Universiteit Utrecht, Utrecht, the Netherlands); P.G.J. Burney (National Heart and Lung Institute, Imperial College, London, UK); R.B. Dantes (Philippine College of Chest Physicians, Manila, Philippines); W. Dejsomritrutai (Dept of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand); D. Dockery (Dept of Environmental Health and Dept 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); E. Falaschetti (Health Survey for England 1995–1996; International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College, London, 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 and the SAPALDIA cohort study (Division of Pulmonary Medicine, University Hospitals of Geneva, Geneva, Switzerland); T. Gislason (Landspítali University Hospital, Dept of Allergy, Respiratory Medicine and Sleep, Reykjavik, Iceland); M. Golshan (Bamdad Respiratory Research Institute, Isfahan, Iran); C.J. Gore (Physiology Dept, Australian Institute of Sport, Belconnen,

Australia); A. Gulsvik (Dept 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 (NHANES, NHANES III Special data sets; Hankinson Consulting, Valdosta, GA, USA); A.J. Henderson (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); M.S.M. Ip (Dept of Medicine, The University of Hong Kong, Pokfulam, Hong Kong, China); C. Janson (Dept of Medical Sciences, Respiratory Medicine and 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 and KORA study (Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Hospital of the Ludwig-Maximilians-University, Munich, Germany); G.S. Kerby (Lung Function Measures in Preschool Children with Cystic Fibrosis study group; University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, CO, USA); J. Kühr (Klinik für Kinder- und Jugendmedizin, Städtisches Klinikum Karlsruhe, Karlsruhe, Germany); S. Kuster and LuftiBus study (Lungenliga Zürich, Zürich, Switzerland); A. Langhammer and The HUNT Study (HUNT Research Centre, NTNU, Verdal, Norway); S. Lum (Portex Respiratory Unit, UCL, Institute of Child Health, London, UK); D.M. Mannino (University of Kentucky, Lexington, Kentucky, USA); G. Marks (Woolcock Institute of Medical Research, Sydney, Australia); A. Miller (Beth Israel Medical Center, New York, NY, USA); G. Mustafa (Dept of Paediatrics, Nishtar Medical College, Multan, Pakistan); E. Nizankowska-Mogilnicka (Division of Pulmonary Diseases, Dept of Medicine, Jagiellonian University School of Medicine, Krakow, Poland); W. Nystad (Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway); Y-M. Oh and Korean NHANES (Dept of Pulmonary and Critical Care Medicine, Asthma Center, Clinical Research Center for Chronic Obstructive Airway Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea); W-H. Pan (Institute of Medical Sciences, Academia Sinica, Taipei, Taiwan); R. Pérez-Padilla and PLATINO study (Instituto Nacional de Enfermedades Respiratorias, Mexico DF, Mexico); P. Piccioni (SC Pneumologia CPA ASL Torino 2, Turin, Italy); F. Pistelli (Pulmonary and Respiratory Pathophysiology Unit, Cardiothoracic Dept, University Hospital of Pisa and Pulmonary Environmental Epidemiology Unit, CNR Institute of Clinical Physiology, Pisa, Italy); K.V.V. Prasad (Government Vemana Yoga Research Institute, Ameerpet, Hyderabad, India); P.H. Quanjer (Dept of Pulmonary Diseases and Dept of Paediatrics, Erasmus Medical Centre, Erasmus University, Rotterdam, the Netherlands); M. Rosenthal (Royal Brompton Hospital, London, UK); H. Schulz and KORA study (Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany); S. Stanojevic (Asthma UK Growing Lungs Initiative; Portex Respiratory Unit, UCL Institute of Child Health, London, UK, and Child Health Evaluative Sciences and Respiratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada); J.B. Soriano and Framingham study (Program of Epidemiology and Clinical Research, CIMERA, Recinte Hospital Joan March, Illes Balears, Spain); W.C. Tan (iCapture Center for Cardiovascular and Pulmonary Research, University of British Columbia, Vancouver, BC, Canada); W. Tomalak (Dept of Physiopathology of Respiratory System, National Institute for TB and Lung Disease, Rabka Branch, Rabka, Poland); S.W. Turner and the SEATON study group (Dept of Child Health, University of Aberdeen, Aberdeen, UK); D. Vilozni (Paediatric Pulmonary Units of The Edmond and Lili Safra Children's Hospital, Sheba Medical Center Ramat-Gan, affiliated with the Sackler Medical School, Tel-Aviv University, Tel-Aviv, Israel); H. Vlachos (Dept of Paediatrics, Division of Respiratory Medicine, University of Sherbrooke, Quebec, QC, Canada); S. West (Respiratory Function Laboratory, Westmead Hospital, Westmead, Australia); E.F.M. Wouters (Maastricht University Medical Center, Maastricht, the Netherlands); Y. Wu (Dept of Occupational Health, School of Public Health, Harbin Medical

University, Harbin, China); D. Zagami (Lung Function Laboratory, Gold Coast Hospital, Southport, Australia); Z. Zhang (Dept of Occupational Health, School of Public Health, Harbin Medical University, Harbin, China); and J.P. Zheng (Guangzhou Institute of Respiratory Disease, State Key Laboratory of Respiratory Disease, Guangzhou, China).

REFERENCES

- 1 Global Lung Function Initiative. Lung Function in Growth and Aging. www.lungfunction.org Date last updated: September 5, 2012. Date last accessed: March 24, 2012.
- 2 NIH Policy on reporting race and ethnicity data: subjects in Clinical Research. <http://grants.nih.gov/grants/guide/notice-files/not-od-01-053.html> Date last updated: June 6, 2006. Date last accessed: March 24, 2012.
- 3 Schrader PC, Quanjer PH, van Zomeren BC, *et al.* Changes in the FEV₁-height relationship during pubertal growth. *Bull Eur Physiopathol Respir* 1984; 20: 381-388.
- 4 Degroodt EG, Quanjer PH, Wise ME, *et al.* Changing relationships between stature and lung volumes during puberty. *Respir Physiol* 1986; 65: 139-153.
- 5 Sherrill DL, Morgan WJ, Taussig LM, *et al.* A mathematical procedure for estimating the spatial relationships between lung function, somatic growth, and maturation. *Pediatr Res* 1989; 25: 316-321.
- 6 Wang X, Dockery DW, Wypij D, *et al.* Pulmonary function between 6 and 18 years of age. *Pediatr Pulmonol* 1993; 15: 75-88.
- 7 Polgar G, Promadhat V, eds. Pulmonary Function Testing in Children: Techniques and Standards. Philadelphia, Saunders, 1971.
- 8 Quanjer PH, Tammeling GJ, Cotes JE, *et al.* Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J* 1993; 6: Suppl. 16, s5-s40.
- 9 Quanjer PH, Borsboom GJJM, Brunekreef B, *et al.* Spirometric reference values for white European children and adolescents: Polgar revisited. *Pediatr Pulmonol* 1995; 19: 135-142.
- 10 Pérez-Padilla R, Valdivia G, Muiño A, *et al.* Spirometric reference values in 5 large Latin American cities for subjects aged 40 years or over. *Arch Bronconeumol* 2006; 42: 317-325.
- 11 Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with discussion). *Appl Statist* 2005; 54: 507-554.
- 12 Stanojevic S, Wade A, Stocks J, *et al.* Reference ranges for spirometry across all ages. A new approach. *Am J Respir Crit Care Med* 2008; 177: 253-260.
- 13 Stanojevic S, Wade A, Cole TJ, *et al.* Spirometry centile charts for young Caucasian children: the Asthma UK Collaborative Initiative. *Am J Respir Crit Care Med* 2009; 180: 547-552.
- 14 Quanjer PH, Stocks J, Cole TJ, *et al.* Influence of secular trends and sample size on reference equations for lung function tests. *Eur Respir J* 2011; 37: 658-664.
- 15 Quanjer PH, Stanojevic S, Stocks J, *et al.* Changes in the FEV₁/FVC ratio during childhood and adolescence: an intercontinental study. *Eur Respir J* 2010; 36: 1391-1399.
- 16 Pellegrino R, Viegi G, Brusasco V, *et al.* Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948-968.
- 17 Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1999; 159: 179-187.
- 18 Global Lung Function Initiative. Lung Function in Growth and Aging. Statutes. www.lungfunction.org/statutes.html Date last updated: September 5, 2012. Date last accessed: March 24, 2012.

- 19 Cole TJ, Stanojevic S, Stocks J, *et al.* Age- and size related reference ranges: a case study of spirometry through childhood and adulthood. *Statist Med* 2009; 28: 880–898.
- 20 van Buuren S, Fredriks M. Worm plot: a simple diagnostic device for modelling growth reference curves. *Statist Med* 2001; 20: 1259–1277.
- 21 Pérez-Padilla R, Regalado-Pineda J, Rojas M, *et al.* Spirometric function in children of Mexico City compared to Mexican-American children. *Pediatr Pulmonol* 2003; 35: 177–183.
- 22 Ip MSM, Karlberg EM, Karlberg JPE, *et al.* Lung function reference values in Chinese children and adolescents in Hong Kong. I. Spirometric values and comparison with other populations. *Am J Respir Crit Care Med* 2000; 162: 424–429.
- 23 Ip MS, Ko FW, Lau AC, *et al.* Updated spirometric reference values for adult Chinese in Hong Kong and implications on clinical utilization. *Chest* 2006; 129: 384–392.
- 24 Pan WH, Chen JY, Haung SL, *et al.* Reference spirometric values in healthy Chinese never-smokers in two townships of Taiwan. *Chin J Physiol* 1997; 40: 165–174.
- 25 Dejsomritrutai W, Nana A, Maranetra KN, *et al.* Reference spirometric values for healthy lifetime nonsmokers in Thailand. *J Med Assoc Thai* 2000; 83: 457–466.
- 26 Jung Keun Choi, Domyung Paek, Jeoung Oh Lee. Normal predictive values of spirometry in Korean population. *Tuberc Respir Dis* 2005; 58: 230–242.
- 27 Wu Y, Zhang Z, Gang B, *et al.* Predictive equations for lung function based on a large occupational population in North China. *J Occup Health* 2009; 51: 417–477.
- 28 Zhang QL, Zheng JP, Yuan BT, *et al.* Feasibility and predicted equations of spirometry in Shenzhen preschool children. *Zhonghua Er Ke Za Zhi (Chinese J Pediatr)* 2005; 43: 843–848.
- 29 National Health and Nutrition Examination Survey IV. National Center for Health Statistics. Data Dissemination Branch, Hyattsville, MD, USA.
- 30 Golshan M, Nematbakhsh M, Amra B, *et al.* Spirometric reference values in a large Middle Eastern population. *Eur Respir J* 2003; 22: 529–534.
- 31 Bougrida M, Ben Saad H, Kheireddinne Bourahli M, *et al.* Équations de références spirométriques des Algériens âgés de 19 à 73 ans [Spirometric reference equations for Algerians aged 19 to 73 years]. *Rev Mal Respir* 2008; 25: 577–590.
- 32 Amra B, Asadi M, Salehi H, *et al.* Normative reference values for lung transfer factor in Isfahan, Iran. *Respirology* 2006; 11: 477–481.
- 33 Trabelsi Y, Ben Saad H, Tabka Z, *et al.* Spirometric reference values in Tunisian children. *Respiration* 2004; 71: 511–518.
- 34 Trabelsi Y, Pariès J, Harrabi I, *et al.* Factors affecting the development of lung function in Tunisian children. *Am J Hum Biol* 2008; 20: 716–725.
- 35 Ben Saad H, Tiftha M, Harabi I, *et al.* Factors influencing ventilatory variables of Tunisian women aged 45 years and more. *Rev Mal Respir* 2006; 23: 324–338.
- 36 Ben Saad H, Rouatbi S, Raoudha S, *et al.* Capacité vitale et débits maximaux expiratoires dans une population nord-africaine âgée de plus de 60 ans. Influence des données anthropométriques et de la parité [Vital capacity and peak expiratory flow rates in a North-African population aged 60 years and over: influence of anthropometric data and parity]. *Rev Mal Respir* 2003; 20: 521–530.
- 37 Trabelsi Y, Tabka Z, Richalet JP, *et al.* Spirometric values in Tunisian children: relationship with pubertal status. *Ann Hum Biol* 2007; 34: 195–205.
- 38 Raju PS, Prasad KVV, Venkata Ramana Y, *et al.* Pulmonary function tests in Indian girls – prediction equations. *Indian J Pediatr* 2004; 71: 893–897.
- 39 Raju PS, Prasad KV, Ramana YV, *et al.* Study on lung function tests and prediction equations in Indian male children. *Indian Pediatr* 2003; 40: 705–711.
- 40 Raju PS, Prasad KVV, Venkata Ramana Y, *et al.* Influence of socioeconomic status on lung function and prediction equations in Indian children. *Pediatr Pulmonol* 2005; 39: 528–536.
- 41 Fulambarker A, Copur AS, Javeri A, *et al.* Reference values for pulmonary function in Asian Indians living in the United States. *Chest* 2004; 126: 1225–1233.
- 42 Al-Riyami BM, Al-Rawas OA, Hassan MO. Normal spirometric reference values for Omani children and adolescents. *Respirology* 2004; 9: 387–391.
- 43 Al-Rawas OA, Baddar S, Al-Maniri AA, *et al.* Normal spirometric reference values for Omani adults. *Lung* 2009; 187: 245–251.
- 44 Yamaguchi-Kabata Y, Nakazono K, Takahashi A, *et al.* Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am J Hum Genet* 2008; 83: 445–456.
- 45 Omoto K, Saitou N. Genetic origins of the Japanese: a partial support for the dual structure hypothesis. *Am J Phys Anthropol* 1997; 102: 437–446.
- 46 Kiefer EM, Hankinson JL, Barr RG. Similar relation of age and height to lung function among Whites, African Americans, and Hispanics. *Am J Epidemiol* 2011; 173: 376–387.
- 47 Quanjer PH, Hall GL, Stanojevic S, *et al.* Age- and height-based prediction bias in spirometry reference equations. *Eur Respir J* 2012; 40: 190–197.
- 48 Brändli O, Schindler Ch, Künzli N, *et al.* Lung function in healthy never smoking adults: reference values and lower limits of normal of a Swiss population. *Thorax* 1996; 51: 277–283.
- 49 Falaschetti E, Laiho J, Primatesta P, *et al.* Prediction equations for normal and low lung function from the Health Survey for England. *Eur Respir J* 2004; 23: 456–463.
- 50 Kuster SP, Kuster D, Schindler C, *et al.* Reference equations for lung function screening of healthy never smoking adults aged 18–80 years. *Eur Respir J* 2008; 31: 860–868.
- 51 Glindmeyer HW, Diem JE, Jones RN, *et al.* Non-comparability of longitudinally and cross-sectionally determined annual change in spirometry. *Am Rev Respir Dis* 1982; 125: 544–548.
- 52 Burrows B, Lebowitz MD, Camilli AE, *et al.* Longitudinal changes in forced expiratory volume in one second in adults: methodologic considerations and findings in healthy nonsmokers. *Am Rev Respir Dis* 1986; 133: 974–980.
- 53 Jedrychowski W, Krzyzanowski M, Wysocki M. Changes in lung function determined longitudinally compared with decline assessed cross-sectionally. The Cracow study. *Eur J Epidemiol* 1986; 2: 134–138.
- 54 Vollmer WM, Johnson LR, McCamant LE, *et al.* Longitudinal versus cross-sectional estimation of lung function decline—further insights. *Stat Med* 1988; 7: 685–696.
- 55 Ware JH, Dockery DW, Louis TH, *et al.* Longitudinal and cross-sectional estimates of pulmonary function decline in never-smoking adults. *Am J Epidemiol* 1990; 132: 685–700.
- 56 Vollmer WM. Reconciling cross-sectional with longitudinal observations on annual decline. *Occup Med* 1993; 8: 339–351.
- 57 Van Pelt W, Borsboom GJJM, Rijcken B, *et al.* Discrepancies between longitudinal and cross-sectional change in ventilatory function in 12 years of follow-up. *Am J Respir Crit Care Med* 1994; 149: 1218–1226.
- 58 Xu X, Laird N, Dockery DW, *et al.* Age, period, and cohort effects on pulmonary function in a 24-year longitudinal study. *Am J Epidemiol* 1995; 141: 554–566.
- 59 Wang ML, McCabe L, Hankinson JL, *et al.* Longitudinal and cross-sectional analyses of lung function in steelworkers. *Am J Respir Crit Care Med* 1996; 153: 1907–1913.
- 60 Kerstjens HAM, Rijcken B, Schouten JP, *et al.* Decline of FEV1 by age and smoking status: facts, figures and fallacies. *Thorax* 1997; 52: 820–827.

- 61 Wang M-L, Petsonk EL. Repeated measures of FEV₁ over six to twelve months: what change is abnormal? *J Occup Environ Med* 2004; 46: 591–595.
- 62 Hnizdo E, Sircar K, Yan T, *et al.* Limits of longitudinal decline for the interpretation of annual changes in FEV₁ in individuals. *Occup Environ Med* 2007; 64: 701–707.
- 63 Quanjer PH, Borsboom GJJM, Kivastik J, *et al.* Cross-sectional and longitudinal spirometry in children and adolescents. Interpretative strategies. *Am J Respir Crit Care Med* 2008; 178: 1262–1270.
- 64 Cotes JE, Gilson JC. Effect of inactivity, weight gain and antitubercular chemotherapy upon lung function in working coal-miners. *Ann Occup Hyg* 1967; 10: 327–335.
- 65 Bande J, Clément J, van de Woestijne KP. The influence of smoking habits and body weight on vital capacity and FEV₁ in male airforce personnel: a longitudinal and cross-sectional analysis. *Am Rev Respir Dis* 1980; 122: 781–790.
- 66 Chen Y, Horne SL, Dosman JA. Body weight and weight gain related to pulmonary function decline in adults: a six year follow up study. *Thorax* 1993; 48: 375–380.
- 67 Chinn DJ, Cotes JE, Reed JW. Longitudinal effects of change in body mass on measurements of ventilatory capacity. *Thorax* 1996; 51: 699–704.
- 68 Beverley AC, Blizzard CL, Schmidt MD, *et al.* Longitudinal associations of adiposity with adult lung function in the Childhood Determinants of Adult Health (CDAH) study. *Obesity* 2011; 19: 2069–2075.
- 69 Parker JM, Dillard TA, Phillips YY. Impact of using stated instead of measured height upon screening spirometry. *Am J Respir Crit Care Med* 1994; 150: 1705–1708.
- 70 Brener ND, Memanus T, Galuska DA, *et al.* Reliability and validity of self-reported height and weight among high school students. *J Adolesc Health* 2003; 32: 281–287.
- 71 Braziniene I, Wilson TA, Lane AH. Accuracy of self-reported height measurements in parents and its effect on mid-parental target height calculation. *BMC Endocrine Disorders* 2007; 7: 2.
- 72 Jansen W, van de Looij-Jansen PM, Ferreira I, *et al.* Differences in measured and self-reported height and weight in Dutch adolescents. *Ann Nutr Metab* 2006; 50: 339–346.
- 73 Lim LLY, Seubsman S-A, Sleight A. Validity of self-reported weight, height, and body mass index among university students in Thailand: implications for population studies of obesity in developing countries. *Population Health Metrics* 2009; 7: 15.
- 74 Wada K, Tamakoshi K, Tsunekawa T, *et al.* Validity of self-reported height and weight in a Japanese workplace population. *Intern J Obesity* 2005; 29: 1093–1099.
- 75 Connor Gorber S, Shields M, Tremblay M. Methodological issues in anthropometry: Self-reported *versus* measured height and weight. Proceedings of Statistics Canada Symposium, 2008.
- 76 Rossiter CE, Weill H. Ethnic differences in lung function: evidence for proportional differences. *Int J Epidemiol* 1974; 3: 55–61.
- 77 Sobol BJ. Assessment of ventilatory abnormality in the asymptomatic subject: an exercise in futility. *Thorax* 1966; 2: 445–449.
- 78 Sobol BJ, Sobol PG. Percent of predicted as the limit of normal in pulmonary function testing: a statistically valid approach. *Thorax* 1979; 34: 1–3.
- 79 Miller MR, Pincock AC. Predicted values: how should we use them? *Thorax* 1988; 43: 265–267.
- 80 Pellegrino R, Viegi G, Brusasco V, *et al.* Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.
- 81 Miller MR, Quanjer PH, Swanney MP, *et al.* Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2011; 139: 52–59.
- 82 Rosenthal M, Bain SH, Cramer D, *et al.* Lung function in white children aged 4–19 years: I – Spirometry. *Thorax* 1993; 48: 794–802.
- 83 Zapletal A, Paul T, Samanek N. Die Bedeutung heutiger Methoden der Lungen-funktionsdiagnostik zur Feststellung einer Obstruktion der Atemwege bei Kindern und Jugendlichen [Significance of contemporary methods of lung function testing for the detection of airway obstruction in children and adolescents]. *Z Erkr Atmungsorgane* 1977; 149: 343–371.
- 84 Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS recommendations. *Am Rev Respir Dis* 1981; 123: 659–664.
- 85 Knudson RJ, Lebowitz MD, Holberg CJ, *et al.* Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis* 1983; 127: 725–734.
- 86 Perkins JM, Khan KT, Smith GD, *et al.* Patterns and trends of adult height in India in 2005–2006. *Econ Hum Biol* 2011; 9: 184–193.
- 87 Gautam RK, Adak DK. Nutrition and Genetic Variation among Central Indian Tribes. In: Proceedings National Symposium of Tribal Health. Jabalpur, Regional Medical Research Centre for Tribals, Indian Council of Medical Research, 2007; pp. 141–153.
- 88 <http://nsdl.niscair.res.in/bitstream/123456789/339/1/pdf+4.4+NISCAIR-Racial-Ethnic-Religious-Linguistic-Groups-India-Text-Revised.pdf> Date last updated: October 21, 2007. Date last accessed: March 24, 2012.
- 89 Malina RM, Brown KH, Zavaleta AN. Relative lower extremity length in Mexican American and in American black and white youth. *Am J Phys Anthropol* 1987; 72: 89–94.
- 90 Hsi BP, Hsu KH, Jenkins DE. Ventilatory functions of normal children and young adults: Mexican-American, white, and black. III. Sitting height as a predictor. *J Pediatr* 1983; 102: 860–865.
- 91 Maca-Meyer N, González AM, Larruga JM, *et al.* Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics* 2001; 2: 13.
- 92 Zheng JP, Zhong N. Normative values of pulmonary function testing in Chinese adults. *Chin Med J* 2002; 115: 50–54.
- 93 Etlar DA. Recent developments in the study of human biology in China: a review. *Hum Biol* 1992; 64: 567–585.
- 94 Zhang HG, Chen YF, Ding M, *et al.* Dermatoglyphics from all Chinese ethnic groups reveal geographic patterning. *PLoS One* 2010; 20, 5: e8783.
- 95 Crapo RO, Jensen RL, Oyunchimeg M, *et al.* Differences in spirometry reference values: a statistical comparison of a Mongolian and a Caucasian study. *Eur Respir J* 1999; 13: 606–609.
- 96 Morgan SL. Richer and taller: stature and living standards in China, 1979–1985. *China J* 2000; 44: 1–39.
- 97 Morgan SL. Stature and Famine in China: The Welfare of the Survivors of the Great Leap Forward Famine, 1959–61, <http://ehsan.econ.usyd.edu.au/papers/Morgan.pdf> Date last accessed: March 24, 2012.
- 98 Morgan SL. Biological indicators of change in the standard of living in China during the 20th century. In: Komlos J, Baten, J, eds. Studies on the Biological Standard of Living in Comparative Perspective. Stuttgart, Franz Steiner Verlag, 1998; pp. 7–34.
- 99 Lin W-S, Zhu F-C, Chen ACN, *et al.* Physical growth of Chinese school children 7–18 years, in 1985. *Ann Hum Biol* 1992; 19: 41–55.
- 100 Ashizawa K, Tanamachi N, Kato S, *et al.* Growth of height and leg length of children of Beijing and Xilinhot, China. *Anthropol Sci* 2008; 116: 67–76.
- 101 Weitz CA, Garruto RM, Chin C, *et al.* Morphological growth and thorax dimensions among Tibetan compared to Han children, adolescents and young adults born and raised at high altitude. *Ann Hum Biol* 2004; 31: 292–310.
- 102 Tanner JM, Hayashi T, Preece MA, *et al.* Increase in length of leg relative to trunk in Japanese children and adults from 1957 to 1977: comparison with British and with Japanese Americans. *Ann Hum Biol* 1982; 9: 411–423.

- 103 Leung SSF, Lau JTF, Xu YY, *et al.* Secular changes in standing height, sitting height and sexual maturation of Chinese – the Hong Kong growth study, 1993. *Ann Hum Biol* 1996; 23: 297–306.
- 104 Ministry of Knowledge Economy. Korean Agency for Technology and Standards. www.kats.go.kr/english/home/home.asp?OlapCode=ATSU15.
- 105 Swanney MP, Jensen RL, Crichton DA, *et al.* FEV₆ is an acceptable surrogate for FVC in the spirometric diagnosis of airway obstruction and restriction. *Am J Respir Crit Care Med* 2000; 162: 917–919.
- 106 Thoracic Society of Australia and New Zealand. Accreditation. www.thoracic.org.au/accreditation Date last accessed: March 24, 2012.
- 107 Standardization of spirometry, 1994 update. American Thoracic Society. *Am J Respir Crit Care Med* 1995; 152: 1107–1136.
- 108 Tanner JM. Growth as a target-seeking function: catch-up and -down growth in man. In: Falkner F, Tanner JM, eds. *Human Growth: A Comprehensive Treatise*. Vol. 2. New York, Plenum Press 1986; pp. 171–209.
- 109 Cole TJ. The secular trend in human physical growth: a biological view. *Econ Hum Biol* 2003; 1: 161–168.
- 110 Sitarama Raju P, Prasad KVV, Venkata Ramana Y, *et al.* Influence of socioeconomic status on lung function and prediction equations in Indian children. *Pediatr Pulmonol* 2005; 39: 528–536.
- 111 Harik-Khan RFI, Muller DC, Wise RA. Racial difference in lung function in African-American and white children: effect of anthropometric, socioeconomic, nutritional, and environmental factors. *Am J Respir Epidemiol* 2004; 160: 893–900.
- 112 Whitrow MJ, Harding S. Ethnic differences in adolescent lung function. Anthropometric, socioeconomic, and psychological factors. *Am J Respir Crit Care Med* 2008; 177: 1262–1267.
- 113 Wu T, Mendola P, Buck GM. Ethnic differences in the presence of secondary sex characteristics and menarche among US girls: the Third National Health and Nutrition Examination Survey, 1988–1994. *Pediatrics* 2002; 110: 752–757.
- 114 Hsi BP, Hsu KHK, Jenkins DE. Ventilatory functions of normal children and young adults: Mexican-American, white, and black. III. Sitting height as a predictor. *J Paediatr* 1983; 102: 860–865.
- 115 Korotzer B, Ong S, Hansen JE. Ethnic differences in pulmonary function in healthy nonsmoking Asian-Americans and European-Americans. *Am J Respir Crit Care Med* 2000; 161: 1101–1108.
- 116 Donnelly PM, Yang T-S, Peat JK, *et al.* What factors explain racial differences in lung volumes? *Eur Respir J* 1991; 3: 829–838.
- 117 White NW, Hanley JH, Lalloo UG, *et al.* Review and analysis of variation between spirometric values reported in 29 studies of healthy African adults. *Am J Respir Crit Care Med* 1994; 150: 348–355.
- 118 Yang T-S, Peat J, Keena V, *et al.* A review of the racial differences in the lung function of normal Caucasian, Chinese and Indian subjects. *Eur Respir J* 1991; 4: 872–880.
- 119 Greksa LP, Spielvogel H, Paz-Zamora M, *et al.* Effect of altitude on the lung function of high altitude residents of European ancestry. *Am J Phys Anthropol* 2005; 75: 77–85.
- 120 Burchard EG, Ziv E, Coyle N, *et al.* The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med* 2003; 348: 1170–1175.
- 121 Cooper RS, Kaufman JS, Ward R. Race and genomics. *N Engl J Med* 2003; 348: 1166–1170.
- 122 Risch N, Burchard E, Ziv E, *et al.* Categorization of humans in biomedical research: genes, race and disease. *Genome Biol* 2002; 3: 7.
- 123 Kumar R, Seibold MA, Aldrich MC, *et al.* Genetic ancestry in lung-function predictions. *N Engl J Med* 2010; 363: 321–330.
- 124 Parra EJ, Kittles RA, Argyropoulos G, *et al.* Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol* 2001; 114: 18–29.
- 125 Jorde LB, Wooding SP. Genetic variation, classification and “race”. *Nat Genet* 2004; 36: Suppl. 11, S28–S33.
- 126 Yaeger R, Avila-Bront A, Abdul K, *et al.* Comparing genetic ancestry and self-described race in African Americans born in the United States and in Africa. *Cancer Epidemiol Biomarkers Prev* 2008; 17: 1329–1338.
- 127 Choudhry S, Taub M, Mei Rui, *et al.* Genome-wide screen for asthma in Puerto Ricans: evidence for association with 5q23 region. *Hum Genet* 2008; 123: 455–468.
- 128 Halder I, Yang BZ, Kranzler HR, *et al.* Measurement of admixture proportions and description of admixture structure in different U.S. populations. *Hum Mutat* 2009; 30: 1299–309.
- 129 Lai CQ, Tucker KL, Choudhry S, *et al.* Population admixture associated with disease prevalence in the Boston Puerto Rican health study. *Hum Genet* 2009; 125: 199–209.
- 130 Rotimi CN, Jorde LB. Ancestry and disease in the age of genomic medicine. *N Engl J Med* 2010; 363: 1551–1558.
- 131 Fulambarker A, Copur AS, Cohen ME, *et al.* Comparison of pulmonary function in immigrant vs US-born Asian Indians. *Chest* 2010; 137: 1398–1404.
- 132 Lung function testing: selection of reference values and interpretative strategies. American Thoracic Society. *Am Rev Respir Dis* 1991; 144: 1202–1218.
- 133 Levy ML, Quanjer PH, Booker R, *et al.* Diagnostic spirometry in primary care. Proposed standards for general practice compliant with American Thoracic Society and European Respiratory Society recommendations. *Prim Care Respir J* 2009; 18: 130–147.
- 134 Corey PN, Ashley MJ, Chan-Yeung M. Racial differences in lung function: search for proportional relationships. *J Occup Med* 1979; 21: 395–398.