

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

Report of the Global Lung Function Initiative (GLI), ERS Task Force to establish improved Lung Function Reference Values.

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

- Philip H. Quanjer Department of Pulmonary Diseases and Department of Paediatrics, Erasmus Medical Centre, Erasmus University, Rotterdam, the Netherlands
- Sanja Stanojevic Portex Respiratory Unit, UCL Institute of Child Health, London, UK and Child Health Evaluative Sciences & Respiratory Medicine, The Hospital for Sick Children, Toronto, Canada
- Tim J. Cole MRC Centre of Epidemiology for Child Health, UCL Institute of Child Health, London, UK
- Xaver Baur Universitätsklinikum Hamburg-Eppendorf, Zentralinstitut für Arbeitsmedizin und Maritime Medizin, Hamburg, Germany
- Graham 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
- Bruce H. Culver Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of Washington, Seattle, Washington, USA
- Paul L. Enright Division of Public Health Sciences, The University of Arizona, Tucson, Arizona, USA
- John L. Hankinson Hankinson Consulting, Athens, Georgia, USA
- Mary S.M. Ip Department of Medicine, The University of Hong Kong, Queen Mary Hospital, Hong Kong, China
- Jinping Zheng Guangzhou Institute of Respiratory Disease, State Key Laboratory of Respiratory Disease, Guangzhou, China
- Janet Stocks Portex Respiratory Unit, UCL Institute of Child Health, London, UK

Corresponding author:
Philip Quanjer, e-mail: pquanjer@xs4all.nl

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 Respirology (APSR); Thoracic Society of Australia and New Zealand (TSANZ) and the American College of Chest Physicians (ACCP).

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

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

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

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

Statistical reviewer: C. Schindler

Persons and centres contributing data to this manuscript:

O.A. *Al-Rawas*, Department of Medicine, College of Medicine and Health, Sciences, Sultan Qaboos University, Muscat, Sultanate of Oman; H.G.M. *Arets*; Department of Pediatric Pulmonology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands; C. *Bárbara*, The Portuguese Society of Pneumology, Lisbon, Portugal; R.G. *Barr*, Columbia University Medical Center, New York, NY, USA and the MESA study; E. *Bateman*, University of Cape Town Lung Institute, Cape Town, South Africa; C.S. *Beardsmore*, Department of Infection, Immunity and Inflammation (Child Health), University of Leicester, Leicester, UK; H. *Ben Saad*, Laboratory of Physiology, Faculty of Medicine, Sousse, University of Sousse, Tunisia; B. *Brunekreef*, Institute for Risk Assessment Sciences, Universiteit Utrecht, Utrecht, the Netherlands; P.G.J. *Burney*, National Heart and Lung Institute, Imperial College, London; R.B. *Dantes*, Philippine College of Chest Physicians, Manila, Philippines; W. *Dejsomritrutai*, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand; D. *Dockery*, Department of Environmental Health, Department of Epidemiology, Boston, MA, USA; H. *Eigen*, Section of Pulmonology and Intensive Care, James Whitcomb Riley Hospital for Children, Indiana University School of Medicine, Indianapolis, IN, USA; E. *Falaszchetti*, [Health Survey for England 1995-1996 (HSE)], International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College, UK; B. *Fallon*, Respiratory Laboratory, Nepean Hospital, Penrith, Australia; A. *Fulambarker*, Pulmonary Division, Rosalind Franklin University of Medicine and Science, The Chicago Medical School, Chicago, IL, USA; M. *Gappa* [LUNOKID study group], Children's Hospital and Research Institute, Marienhospital Wesel, Germany; M.W. *Gerbase*, Division of Pulmonary Medicine, University Hospitals of Geneva, Geneva, Switzerland, and the SAPALDIA cohort study; T. *Gislason*, Landspítali University Hospital, Dept. of Allergy, Respiratory Medicine and Sleep, Reykjavik, Iceland; M. *Golshan*, Bamdad Respiratory Research Institute, Isfahan, Iran; C.J. *Gore*, Physiology Department, Australian Institute of Sport,

Belconnen, Australia; A. *Gulsvik*, Department of Thoracic Medicine, Institute of Medicine, University of Bergen, Bergen, Norway; G.L. *Hall*, Respiratory Medicine, Princess Margaret Hospital for Children, Perth, Australia; J.L. *Hankinson*, [NHANES, NHANES III Special data sets], Hankinson Consulting, Valdosta, GA, USA; A.J. *Henderson*, [ALSPAC, <http://www.bris.ac.uk/alspac>], University of Bristol, Bristol, UK; E. *Hnizdo*, Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, Morgantown, WV, USA; M.S.M. *Ip*, Dept of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China; C. *Janson*, Department of Medical Sciences: Respiratory Medicine & Allergology, Uppsala University, Sweden; C. *Jenkins*, Woolcock Institute of Medical Research, Sydney, Australia; A. *Jithoo*, University of Cape Town Lung Institute, Cape Town, South Africa; S. *Karrasch*, Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Hospital of the Ludwig-Maximilians-University, Munich, Germany [KORA study]; G.S. *Kerby* (Lung Function Measures in Preschool Children with Cystic Fibrosis study 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*, Lungenliga Zürich, Zürich, Switzerland [LuftiBus study]; A. *Langhammer*, [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*, Department of Pediatrics, Nishtar Medical College, Multan, Pakistan; E. *Nizankowska-Mogilnicka*, Division of Pulmonary Diseases, Department of Medicine, Jagiellonian University School of Medicine, Cracow, Poland; W. *Nystad*, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; Y-M. *Oh* [Korean NHANES], Department 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*, Instituto Nacional de Enfermedades Respiratorias, Mexico DF, Mexico, [PLATINO study]; P. *Piccioni*, SC Pneumologia CPA ASL Torino 2, Torino, Italy; F. *Pistelli*, Pulmonary and Respiratory Pathophysiology Unit, Cardiothoracic Department, University Hospital of Pisa and Pulmonary Environmental Epidemiology Unit, CNR Institute of Clinical Physiology, Pisa, Italy; *Prasad* KVV, Government Vemana Yoga Research Institute, Ameerpet, Hyderabad, India; P.H. *Quanjer*, Department of Pulmonary Diseases, and Department of Pediatrics, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands; M. *Rosenthal*, Royal Brompton Hospital, London, UK; H. *Schulz*, Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany [KORA study]; S. *Stanojevic*, Portex Respiratory Unit,

UCL Institute of Child Health, London, UK, [Asthma UK Growing Lungs Initiative (<http://www.growinglungs.org.uk>)], and Child Health Evaluative Sciences & Respiratory Medicine, The Hospital for Sick Children, Toronto; J.B. *Soriano*, Program of Epidemiology and Clinical Research, CIMERA, Recinte Hospital Joan March, Illes Balears, Spain [Framingham study]; W.C. *Tan*, iCapture Center for Cardiovascular and Pulmonary Research, University of British Columbia, Vancouver, BC, Canada; W. *Tomalak*, Dept. Physiopathology of Respiratory System, National Institute for TBC & Lung Dis., Rabka Branch, Poland; S.W. *Turner* [The SEATON study group], Department of Child Health, University of Aberdeen, Aberdeen, UK; *Vilozni*, Pediatric 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*, Department of Pediatrics, Division of Respiratory Medicine, University of Sherbrooke, Quebec, Canada; S. *West*, Respiratory Function Laboratory, Westmead Hospital, Australia; E.F.M. *Wouters*, Maastricht University Medical Center, Maastricht, the Netherlands; Y. *Wu*, Department of Occupational Health, School of Public Health, Harbin Medical University, Harbin, China; D. *Zagami*, Lung Function Laboratory, Gold Coast Hospital, Southport, QLD, Australia; Z. *Zhang*, Department of Occupational Health, School of Public Health, Harbin Medical University, China; J.P. *Zheng*, Guangzhou Institute of Respiratory Disease, State Key Laboratory of Respiratory Disease, Guangzhou, China.

This study includes data from the MESA study, Funded by National Institutes of Health R01-HL077612, N01-HC095159-165, N01-HC095169.

INTRODUCTION

1.1. Objectives

2. METHODS

2.1. Recruitment and data collection

2.2. Data

2.3. Quality control

2.4. Indices considered

2.5. Statistical analyses

2.6. Prediction models

2.7. Defining groups

3. RESULTS

3.1. Data

3.2. Stature

3.3. Datasets by region

3.3.1. Latin-America

3.3.2. East Asia

3.3.3. African-American data

3.3.4. Mexican-American data

3.3.5. North Africa and Iran

3.3.6. Indian subcontinent

3.3.7. Oman

3.3.8. Japan

3.4. Pooling groups

3.4.1. Other

3.5. Prediction models

3.6. Simplifying look-up tables

3.7. Proportional differences between ethnic groups

3.8. Effect of collation on predicted values and LLN

3.9. Age-related changes in pulmonary function

3.10. Entering height and age into the equations

4. DISCUSSION

4.1. Representativeness of equations for various groups

4.2. Data quality

4.3. Secular trends

4.4. Age related trends in FEV₁ and FVC

4.5. Mixed ethnic descent

4.6. Application in clinical practice

4.7. Clinical decision making

4.8. Strengths and weaknesses

4.9. Future developments

5. CONCLUSIONS

6. RECOMMENDATIONS

References

ABSTRACT

Objective: Derive continuous prediction equations and their lower limits of normal for spirometric indices, which are applicable globally.

Material: 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 years.

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

Results: 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 from 3-95 years for Caucasians (N=57,395), African Americans (N=3,545), and North (N=4,992) and South East Asians (N=8,255). FEV₁ and 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.

Conclusion: 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, Arab, Polynesian, Latin American countries, and Africa will further improve these equations in the future.

1 INTRODUCTION

Pulmonary function tests fulfill a pivotal role in respiratory medicine. They are used to diagnose airways obstruction, assess its severity and prognosis, delineate risk factors (*e.g.*, preoperative 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) 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 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 [7] up to age 18 years, and from the ECSC/ERS [8] for adults, a 175 cm tall boy producing a forced vital capacity (FVC) of 4.2 L is at 97% of predicted at 17.9 years 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 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 non-smokers 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, 11] to collated data from 11 countries; this allowed modelling 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 year period [14], validating the use of collated data. Another benefit of using collated data was the identification of a biphasic trend in the FEV₁/FVC ratio in childhood and adolescence [15].

In 2005, the American Thoracic Society (ATS) and European Respiratory Society (ERS) published recommendations for standardised lung function testing [16]. One set of spirometric reference equations for adults [17], and one for children and adolescents [6], was 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 valid world-wide, applicable to as many ethnic groups as possible. From 2006, an increasing number of centres agreed to share data with one of the authors (PQ); subsequently the Global Lungs Initiative was established in Berlin in September 2008, acquiring ERS Task Force status in April 2010. The Global Lungs Initiative was subsequently endorsed by the ATS, Australian and New Zealand Society of Respiratory Science (ANZSRS), Asian Pacific Society for Respirology (APSR) and the Thoracic Society of Australia and New Zealand (TSANZ).

1.1 Objectives

The objectives of the ERS Global Lungs Initiative are to establish improved international spirometry reference equations that

- are based on individual lung function data collected under standardised measurement conditions with documented equipment and software
- are modelled using modern statistical techniques to allow continuous equations across the entire age range from early childhood to old age
- 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
- are clinically useful and can be incorporated into commercially available equipment
- are reported in such a manner as to give a clear indication of where the subject lies with respect to the 'normal range'.

2 METHODS

2.1 Recruitment and data collection

We contacted individuals, groups or organisations that were known to have collected representative spirometry data from asymptomatic lifelong non-smokers, 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.

2.2 Data

Datasets were obtained from 73 centres (initial N=160,330). 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=3483) 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 years ([Figure 1](#), online supplement (OLS) tables [E1](#), [E2](#) and [E3](#)).

2.3 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 spirograms.

2.4 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 years (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 when 75% of the FVC has been exhaled (FEF₇₅) 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.

2.5 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, sigma) method, using the GAMLSS package [11] in the statistical software R [Version 2.14.1; R Foundation, <http://www.r-project.org>]; this allows modelling 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 were as described by Cole *et al.* [19].

2.6 Prediction models

The LMS method, imbedded in GAMLSS, allows modelling the expected mean (μ or M), the coefficient of variation (σ or S) (CoV), and skewness (λ or L). 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. A document to help guide those who may wish to use GAMLSS, and which explains the procedures with worked examples, can be downloaded from <http://www.spirxpert.com/download/GAMLSS-in-action.zip>.

Applying the methodology described by Cole *et al.* [19] the best fit was estimated using the Box-Cox-Cole-Green (BCCG) distribution. The optimal degrees of freedom (df) for the spline curve was chosen to minimise the Schwarz Bayesian Criterion (SBC), where adding one df to the model penalises the deviance by $\ln(N)$ units, N being the sample size. As N for males and females was $\sim 30,000$ - $40,000$, 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:

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

where Y = dependent variable, H = standing height (cm); A = age (yr); a , b , c , d_1 and d_2 are coefficients which vary for each ethnic group, and spline is an age-specific contribution from the spline function. Group is a dummy variable with values 0 or 1 indicating ethnicity, where Caucasians are the reference. Any of Y , H or A may be log transformed (see [OLS section 2](#) for details). Essentially, we are therefore dealing with a linear regression equation with an age-specific correction in the form of the age-spline.

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.

2.7 Defining groups

Data were available from 5 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.

3 RESULTS

3.1 Data

The total number of subjects originally included in the study (*i.e.* prior to exclusion for reasons listed in section 2.2) was 97,759 (55.3% females), age range 2.5-95 years ([Table 1](#), OLS table [E2](#) and [E3](#));

47.7% were ≤ 20 years, and 0.8% ≥ 80 years old. These data were combined from 72 datasets from 33 countries (OLS Table E1). Representation was poor from South America, and absent from Malaysia, Indonesia and sub-Saharan Africa.

3.2 Stature

Stature is the main determinant of pulmonary function therefore we investigated whether stature differed significantly between populations (see OLS, Figure E2, E3). While there were significant differences in stature-for-age between populations, the variability between groups was minimal. The within-group coefficient of variation (CoV) was largest in preschool children (OLS Figure E3), declining rapidly towards adolescence followed by an increase, reflecting differences in the timing of the pubertal growth spurt. There was then a drop until about age 30, followed by a small but steady increase towards old age. With one exception, the maximum difference in the CoV between populations was about 1%. The CoV was considerably larger in Indian and Pakistani schoolchildren and adolescents than in other populations (OLS Figure E3).

3.3 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, United Kingdom, USA, Uruguay, Venezuela. Not all of these data could be used (OLS Table 1).

3.3.1 Latin-America

Six datasets were available from Latin America: five related to adults (Mexico City, Sao Paulo, Caracas, Montevideo, Santiago) [10], and the remaining one to Mexican children and young adults [21]. The samples differed in height (OLS Figure E4) and predicted spirometric values (OLS Figure E5). 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 years of age.

3.3.2 East Asia

Nine datasets with 13,247 records (OLS Table E3) were available from Hong Kong (China) [22,23], Taiwan [24], Thailand [25], the USA [26], Korea [27], China [28,29], and one from Chinese children aged 3-6 years [30]; all datasets were from urban populations. There were significant differences in

standing height between centres, Chinese in the USA and mainland China being taller, and people in Thailand, Taiwan, Hong Kong and southwest China shorter (OLS Figure E6 and E7).

Regression analysis revealed significant differences for spirometric indices between centres. While there was remarkable agreement between 6 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 [29] and Korea [27] (OLS Figure E6). 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.

3.3.3 African-American data

Four African-American datasets were available [6,17,26,31] 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.

3.3.4 Mexican-American data

Four Mexican-American datasets were available [6,17,26,31] 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 -0.04).

3.3.5 North Africa and Iran

Six datasets [32-39] were available comprising 7,273 subjects (54.1% male) aged 5-90 years. 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 fit into any group. For example, predicted FEV₁/FVC (5th centile) in North African and Iranian women

age 60 years, 168 cm, are 0.79 (0.67) and 0.82 (0.74), respectively; in 60 year old men (180 cm) the corresponding values are 0.78 (0.66) and 0.82 (0.74).

3.3.6 Indian subcontinent

Data on children were available from India [40-42] and Pakistan [43], and on adults of Asian Indian descent from the USA [44] (total N=5,477, [OLS Table E3](#)). 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.

3.3.7 Oman

Data on 1,256 lifelong non-smokers (51.2% males) aged 6-65 years were available [45,46]. They did not fit in any of the 4 groups that were formed, could not be combined with data from Iran, and are therefore not included in the present (Global Lungs 2012) prediction equations ([OLS Table E3](#)).

3.3.8 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 years) has recently been collected as part of the Japan Lung Health Survey [personal communication, 47]. Given the “dual-origin hypothesis”, according to which ancestral Japanese populations were brought by two major pre-historic migration events, people migrating into Japan from the north and south of the Asian continent [48-49], 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 section 3.4.1).

3.4 Pooling groups

We set the criteria for combining groups as: (a) minimal effects on predicted mean values and (b) small effects on the clinically important lower limit of normal (LLN). The latter was checked by inspection of the percentage of subjects in groups alluded to above whose observations fell below the fifth 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 (N<150) sub-sets wherein average z-score deviated by > ±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:

Group	Country/region
'Caucasian'	Europe, Israel, Australia, USA, Canada, Mexican Americans, Brazil, Chile, Mexico, Uruguay, Venezuela, Algeria, Tunisia
Black	African American
South East Asian	Thailand, Taiwan and China (including Hong Kong) south of the Huaihe River and Qinling Mountains
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 [50].

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 [32,34] 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 could therefore not be included. Data on Latin Americans are therefore limited to those over 40 years old. This left 74,187 subjects in 4 groups ([Table 2](#), OLS Tables [E2](#) and [E3](#)).

3.4.1 'Other'

In the reference equations there are two distinct adjustments for each of the four ethnic groups, one relating to the mean (M) and the other to the CoV (S) (see sections 3.5 and 3.7). 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 M and S 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 M and S 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.

3.5 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):

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

where group takes a value 0 or 1 for Caucasians, African Americans, North or South East Asians, as appropriate (OLS section 3.2), and the coefficient d differs between groups. A smoothing age spline was invariably required for predicting the mean (M) and coefficient of variation (S) for FEV₁, FVC, FEV₁/FVC, FEF_{25-75%} and FEF₇₅; it was also required for modelling skewness (L , λ), 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 year age range a simple linear model (*i.e.* without look-up tables) sufficed.

Note that the d coefficients for M and S in each ethnic group correspond to the corresponding adjustments in section 3.4.1 above.

3.6 Simplifying look-up tables

As delineated in sections 2.6 and 3.10 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 years and older with an equation, without loss of accuracy, so that lung function devices with limited memory can significantly improve on the use of resources (see OLS [sections 2 and 3](#)). The 3-25 years 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 years), as a few months age difference can affect the predicted values by up to 8.5% [51]. Look-up tables are available for download as part of the online data supplement to this manuscript.

3.7 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](#)).

3.8 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 [17,31,52-54] were compared with results from the collated dataset. The CoV for these five studies was calculated by expressing the residual standard deviation ($(\text{predicted} - \text{LLN})/1.644$) as a percent of the predicted value. With the exception of NHANES III and NHANES IV [17,31], 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 two ([Figure 3](#), [OLS Figure E8](#)); this has been attributed to the use of different equipment [54], 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. ([Figure 3](#), [OLS Figure E8](#)). Datasets with unusually high or low average predicted values might influence both predicted mean and scatter. As discussed

in section 3.4, removing sets in which the average z-score deviated by $> \pm 0.4$ (i.e. more than $\sim 5.5\%$ difference) from the overall mean, led to trivial changes in predicted values and their LLN.

3.9 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 than in tall people. This is shown in [Figure 4](#) for FEV₁ in adult Caucasian males of height 160, 175 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 ([Figure 4](#)). Please note that longitudinal changes differ from cross-sectional age differences [55-67], and are affected by changes in body weight [68-72].

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

3.10 Entering height and age into the equations

The prediction equations have the form:

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

where a is the intercept, H = height (cm), b the exponent for height, A = age (years) and c the exponent for age, and spline the contribution from the age spline; group is Caucasian, African American, South or North East Asian, and takes the value 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 = 2.22. Entering H = 160 cm instead of H = 161 cm changes the predicted value by 1.3%. The calculation of the error for age 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 years 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 years error in age in a 14.9 year old boy 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 [73-79]. 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 accuracy.

The look-up tables for the spline in age have a resolution of 0.25 years, *i.e.* 3 months. In the above example a 3 months 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 years of age onwards, the look-up table can be replaced by equations which, if used, avoid the need for interpolation and loss of accuracy (see OLS [sections 2](#) and [3](#) for details).

4 DISCUSSION

This is the first study to present spirometry prediction equations spanning ages 3-95 years 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 FEV_1/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 FEV_1 , leading to falls in FEV_1/FVC , a trend that is reversed in adolescence, a pattern that went undetected with conventional statistical techniques [15] (Figure 2). By undertaking these analyses, it was possible to confirm that this is a physiological pattern of development reflecting differential growth of FEV_1 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 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 [80] ([Table 3](#)), signifying proportionate scaling of lung size due to differences in body build, so that the FEV_1/FVC ratio is generally independent of ethnic group. This has clinical advantages in that, with the exception of South East Asians, in whom the FEV_1/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 FEV_1/FVC across ethnic groups. Whereas the FEV_1/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 sub-continent, for whom published literature suggests that predicted values may be at least as low as those found in Black subjects. With on-going data collection, it is hoped that the majority of 'missing groups' will be included within the next 5 years. As discussed in section 3.4.1, provided a sufficiently large and representative dataset is collected using standardised protocols, it 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 percent of the predicted value; however, this approach has important clinical consequences, particularly when used to classify patients and disease severity using a fixed cut-off [81-85]. 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 [12]. This study confirms and extends this observation by showing that it occurs in all ethnic and geographically defined groups (Figure 5), including a temporary increase in variability during the adolescent growth spurt. This pattern was observed in individual studies with large sample size and with known good quality control (Figure 3), further supporting that this is not an artifact.

4.1 Representativeness of equations for various groups

Comparison of predicted values in four large recent studies [17,52-54] 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, OLS Figure E8). 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 lifelong non-smokers 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 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 [42,86,87]. The 1 billion inhabitants of the Indian subcontinent are comprised of a large number of ethnic groups from various racial strains [88] from South to North; socio-economically 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, 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 Hong Kong adults to those of Hankinson (NHANES III), differences in FVC varied between 12 and 17% depending on age and sex [23], similarly to the present findings ([Table 3](#)). Yet 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 socio-economic 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 socio-economic 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.

4.2 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 lower limits of

normal 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 4 large recent studies with good quality control ([Figure 3](#), [OLS Figure E8](#)) [17,52-54].

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], over 30% of measurements needed to be discarded because they did not meet ATS quality criteria [107]. The great majority of these exclusions related to subjects <30 years; the major stumbling block was the requirement that subjects should exhale for at least 6 seconds. 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 seconds. 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.

4.3 Secular trends

In Caucasian subjects, with data collected over 3 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 socio-economic conditions lead to greater physical development and improved health.

4.4 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 than in short people ([Figure 4](#)). For example, in a Caucasian adult male there is an accelerating cross-sectional decline in FEV₁ after age 30 years, with a nadir at age 62 when the annual loss ranges between 32 and 46 mL. The decline then decelerates, possibly reflecting a healthy survivor effect.

4.5 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), socio-economic 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.

4.6 Application in clinical practice

For any given height and sex, a one year difference in age can alter the predicted value by up to 8.5% [51] in those under 20 years of age. 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 [6,7,131,132], 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: the absolute values for FEV₁ and FVC differ by a certain percentage, and 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 ([Figure 2](#)).

Caution is required interpreting the prediction equations for those over age ~75-80 years, where the sample size is small ([Figure 1](#), [OLS Table E5](#)). This also holds for North East Asians under 15 years of age ([Figure 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, nor 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 [133]. Hence more studies are required to elucidate the influence of country of birth on the level of lung function.

4.7 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 ± 2 (strictly 1.96) standard deviations, which extends from the 2.5th to the 97.5th centile of the distribution. The American Thoracic Society and the European Respiratory Society [8,84,134] both recommend the use of the 5th centile to define the lower limit of normal (*i.e.* -1.64 z-scores). z-scores indicate how many standard deviations (SD) a measurement is from its predicted value. Unlike percent of predicted, they are free from bias due to age, height, sex and ethnic group, and are therefore particularly useful in defining the lower and upper limits of normal; they also simplify uniform interpretation of test results, particularly if presented as pictograms ([Figure 8](#)); software with the latter facility is freely available from www.lungfunction.org.

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 the above, 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, a LLN at the 5th centile (z-score -1.64) is clinically acceptable.

It is common practice to regard 80 percent of predicted as the lower limit of normal. However, the true LLN, expressed as a percent of predicted, varies considerably with age ([Figure 6](#) and [7](#)). Hence, the use of a fixed threshold such as 80% percent of predicted for FEV₁ and FVC or 0.7 for the FEV₁/FVC ratio across all ages is discouraged as it leads to significant misclassification [85] ([Figure 7](#)). In the collated data a z-score for FEV₁/FVC below the 5th centile was associated with a FEV₁ below the 5th centile in 1.2% of cases, with FEV₁ percent predicted <80% in 1.3% of cases, and with FEV₁ percent predicted <70% in 0.4% of cases. In those aged over 40 years 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 FEF₇₅ are measured conditional 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 20-62% for FEF_{25-75%} and between 27-89% for FEF₇₅ ([OLS Figure E9](#)). While the FEF_{75%} and FEF_{25-75%} are not among the indices recommended by ECSC/ERS [8], ATS [107,134] 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 coefficients of variation severely limit the use of these indices for diagnostic purposes in adults, but this does not preclude smaller coefficients of variation within subjects, nor the use of these indices in aetiological studies where differences between groups may provide valuable clues. FEV₃ and FEV₆ were not considered since so 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 under age 20 years (20-33% for FEF_{25-75%}, and 27-42% for FEF₇₅) further research is also required to determine whether such outcomes are in fact clinically useful.

4.8 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 for several ethnic groups simultaneously, confirming the presence of proportional differences between them ([Figure 2](#), [Table 3](#)) [80,135]. Results in Caucasians are comparable to those in large recent studies ([OLS Figure E8](#)). Data from Africa, Polynesian peoples, 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; hence, 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 [52,54]; 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 Stanojevic 'all-age' reference equations [13] and are therefore in an excellent position to update to the present equations [136], 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.

4.9 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.

5 CONCLUSIONS

This study has led to the derivation of continuous equations for predicted values and age-appropriate LLN for spirometric indices from age 3-95 years, based on 74,187 records from healthy non-smoking males and females from 26 countries across five continents. Ethnic and geographic groups can be grouped under the headings Caucasian, African American, North and South East Asian. However, information on pulmonary function is still incomplete in many parts of the world, and more data on subjects over age 75 years are required. In the meantime, the predicted values and reference range for those over ~75-80 years in all groups, and in North East Asians below age 15 years, should be interpreted with caution. Comparison with other large studies confirms that the collated data are of good quality, 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.

6 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 between 3–75 years of age, 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 above 75 years of age, and for pre-school children.
- More studies are required in non-Caucasians, particularly Arab, Indian, Polynesian, African and Latin American peoples, 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 between 3-95 years.

- 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 re-calculating 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 lower limit of normal as a fixed percent 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 lower limit of normal (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 a 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, a LLN corresponding to the 2.5th centile (LLN 2.5%, z-score -1.96) is recommended as the decision limit.

Acknowledgement

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

References

1. www.lungfunction.org/publishedreferencevalues.html, last visited 24 March 2012.
2. <http://grants.nih.gov/grants/guide/notice-files/not-od-01-053.html>.
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. Degroot 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, Landau LM, Burrows B. 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, Fay ME, Ferris BG Jr. Pulmonary function between 6 and 18 years of age. *Pediatr Pulmonol* 1993; 15: 75-88.
7. Polgar G, Promadhat V. 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): 5-40.
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, Hall GL, Stanojevic S. 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. ATS/ERS Task Force: Standardisation of lung function testing. 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. www.lungfunction.org/statutes.html, last visited 24 March 2012.
19. Cole TJ, Stanojevic S, Stocks J, Coates AL, Hankinson JL, Wade AM. 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, Luk KDK, Leong JCY. 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 MSM, Ko FWS, ACW Lau, Yu WC, Tang KS, Choo K, Chan-Yeung MMW. 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. MESA Lung Study, courtesy dr G. Barr.
27. Jung Keun Choi, Domyung Paek, Jeoung Oh Lee. Normal predictive values of spirometry in Korean population. *Tuberc Respir Dis* 2005; 58: 230-242.
28. Chinese data, courtesy Prof. Zheng.
29. Wu Y, Zhang Z, Gang B, Love EJ. Predictive equations for lung function based on a large occupational population in North China. *J Occup Health* 2009; 51: 417-477.
30. Zhang QL, Zheng JP, Yuan BT, He H, Wang J, An JY, Zhang M, Luo DF, Chen GL. Feasibility and predicted equations of spirometry in Shenzhen preschool children. *Zhonghua Er Ke Za Zhi (Chinese J Pediatr)* 2005; 43: 843-848.
31. National Health and Nutrition Examination Survey IV. National Center for Health Statistics. Data Dissemination Branch, Hyattsville, MD, USA.
32. Golshan M, Nematbakhsh M, Amra B, Crapo RO. Spirometric reference values in a large Middle Eastern population. *Eur Respir J* 2003; 22:529-534. Authors' correction: *Eur Respir J* 2004; 23: 356

33. Bougrida M, Ben Saad H, Kheireddinne Bourahli M, Bougmiza I, Mehdioui H. Équations de références spirométriques des Algériens âgés de 19 à 73 ans. *Rev Mal Respir* 2008; 25: 577-590.
34. Amra B, Asadi M, Salehi H, Zamani AR, Golshan M. Normative reference values for lung transfer factor in Isfahan, Iran. *Respirology* 2006; 11: 477–481.
35. Trabelsi Y, Ben Saad H, Tabka Z, Garbi N, Bouchez Buvry A, Richalet JP, Guenard H. Spirometric reference values in Tunisian children. *Respiration* 2004; 71: 511-518.
36. Trabelsi Y, Pariès J, Harrabi I, Zbidi A, Tabka Z, Richalet JP, Buvry A. Factors affecting the development of lung function in Tunisian children. *Am J Hum Biol* 2008; 20: 716-725.
37. Ben Saad H, Tfifha M, Harabil, *et al.* Factors influencing ventilatory variables of Tunisian women aged 45 years and more. *Rev Mal Respir* 2006; 23: 324-338.
38. Ben Saad H, Rouatbi S, Raoudha S, Tabka Z, Laouani Kechrid C, Hassen G, Guénard H. 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. *Rev Mal Respir* 2003; 20: 521-530.
39. Trabelsi Y, Tabka Z, Richalet JP, Gharbi N, Bienvenu A, Guénard H, Buvry A. Spirometric values in Tunisian children: relationship with pubertal status. *Ann Hum Biol* 2007; 34: 195-205.
40. Raju PS, Prasad KVV, Venkata Ramana Y, Murthy KJR. Pulmonary function tests in Indian girls – prediction equations. *Indian J Pediatr* 2004; 71: 893-897.
41. Raju PS, Prasad KV, Ramana YV, Ahmed SK, Murthy KJ. Study on lung function tests and prediction equations in Indian male children. *Indian Pediatr* 2003; 40: 705-711.
42. Raju PS, Prasad KVV, Venkata Ramana Y, Murthy KJR. Influence of socioeconomic status on lung function and prediction equations in Indian children. *Pediatr Pulmonol* 2005; 39: 528–536.
43. Dr. Ghulam Mustafa, Pakistani children, unpublished work.
44. Fulambarker A, Copur AS, Javeri A, Jere S, Cohen ME. Reference values for pulmonary function in Asian Indians living in the United States. *Chest* 2004; 126: 1225-1233.
45. Al-Riyami BM, Al-Rawas OA, Hassan MO. Normal spirometric reference values for Omani children and adolescents. *Respirology* 2004; 9: 387–391.
46. Al-Rawas OA, Baddar S, Al-Maniri AA, Balaji J, Jayakrishnan B, Al-Riyami BM. Normal spirometric reference values for Omani adults. *Lung* 2009; 187: 245-251.
47. M. Kubota, H. Kobayashi, K. Tatsumi, H. Omori and Japanese Lung Health Survey Group: a joint project of the Japan Respiratory Society and Japan Society of Ningen Dock, personal communication.

48. Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, Kubo M, Nakamura Y, Kamatani N. 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.
49. 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.
50. 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.
51. Quanjer PH, Hall GL, Stanojevic S, Cole TJ, Stocks J, on behalf of the Global Lungs Initiative. Age- and height-based prediction bias in spirometry reference equations. *ERJ Express*. Published on December 20, 2011 as doi: 10.1183/09031936.00161011.
52. Brändli O, Schindler Ch, Künzli N, Keller R, Perruchoud AP, and SAPALDIA team. Lung function in healthy never smoking adults: reference values and lower limits of normal of a Swiss population. *Thorax* 1996; 51: 277-283.
53. Falaschetti E, Laiho J, Primatesta P, Purdon S: Prediction equations for normal and low lung function from the Health Survey for England. *Eur Respir J* 2004; 23: 456-463.
54. Kuster SP, Kuster D, Schindler C, Rochat MK, Braun J, Held L, Brändli O. Reference equations for lung function screening of healthy never smoking adults aged 18-80 years. *Eur Respir J* 2008; 31: 860-868.
55. Glindmeyer HW, Diem JE, Jones RN, Weill H. Non-comparability of longitudinally and cross-sectionally determined annual change in spirometry. *Am Rev Respir Dis* 1982;125:544–548.
56. Burrows B, Lebowitz MD, Camilli AE, Knudson RJ. 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.
57. 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.
58. Vollmer WM, Johnson LR, McCamant LE, Buist AS. Longitudinal versus cross-sectional estimation of lung function decline--further insights. *Stat Med* 1988; 7: 685-696.
59. Ware JH, Dockery DW, Louis TH, Xu X, Ferris BG Jr JR., Speizer FE. Longitudinal and cross-sectional estimates of pulmonary function decline in never-smoking adults. *Am J Epidemiol* 1990; 132: 685-700.
60. Vollmer WM. Reconciling cross-sectional with longitudinal observations on annual decline. *Occup Med* 1993; 8: 339-351.

61. Van Pelt W, Borsboom GJJM, Rijcken B, Schouten JP, van Zomeren BC, Quanjer PH. 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.
62. Xu X, Laird N, Dockery DW, Schouten JP, Rijcken B, Weiss ST. Age, period, and cohort effects on pulmonary function in a 24-year longitudinal study. *Am J Epidemiol* 1995; 141: 554–566.
63. Wang ML, McCabe L, Hankinson JL, Shamsain MH, Gunel E, Lapp NL, Banks DE. Longitudinal and cross-sectional analyses of lung function in steelworkers. *Am J Respir Crit Care Med* 1996; 153: 1907-1913.
64. Kerstjens HAM, Rijcken B, Schouten JP, Postma DS. Decline of FEV₁ by age and smoking status: facts, figures and fallacies. *Thorax* 1997; 52: 820-827.
65. 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.
66. Hnizdo E, Sircar K, Yan T, Harber P, Fleming J, Glindmeyer HW. Limits of longitudinal decline for the interpretation of annual changes in FEV₁ in individuals. *Occup Environ Med* 2007; 64: 701–707.
67. Quanjer PH, Borsboom GJJM, Kivastik J, Merkus PJFM, Hankinson JL, Houthuijs D, Brunekreef B, Ihorst G, Kühn J. Cross-sectional and longitudinal spirometry in children and adolescents. Interpretative strategies. *Am J Respir Crit Care Med* 2008; 178: 1262–1270.
68. 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.
69. 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.
70. 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.
71. Chinn DJ, Cotes JE, Reed JW. Longitudinal effects of change in body mass on measurements of ventilatory capacity. *Thorax* 1996; 51: 699-704.
72. Beverley AC, Blizzard CL, Schmidt MD, Walters EH, Dwyer T, Venn AJ. Longitudinal associations of adiposity with adult lung function in the Childhood Determinants of Adult Health (CDAH) study. *Obesity* 2011; 19: 2069-2075.
73. 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(6 Pt 1):1705-1708.
74. Brener ND, Mcmanus T, Galuska DA, Lowry R, Wechsler H. Reliability and validity of self-reported height and weight among high school students. *J Adolesc Health* 2003; 32: 281-287.

75. Braziuniene 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.
76. Jansen W, van de Looij-Jansen P. M, Ferreira I, de Wilde EJ, Brug J. Differences in measured and self-reported height and weight in Dutch adolescents. *Ann Nutr Metab* 2006; 50: 339-346.
77. 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.
78. Wada K, Tamakoshi K, Tsunekawa T, Otsuka R, Zhang H, Murata C, Nagasawa N, Matsushita K, Sugiura K, Yatsuya H, Toyoshima H. Validity of self-reported height and weight in a Japanese workplace population. *Intern J Obesity* 2005; 29: 1093–1099.
79. Shields M, Gorber SC, Tremblay MS. Methodological issues in anthropometry: Self-reported versus measured height and weight. Proceedings of Statistics Canada Symposium 2008.
80. Rossiter CE, Weill H. Ethnic differences in lung function: evidence for proportional differences. *Int J Epidemiol* 1974; 3: 55-61.
81. Sobol BJ. Assessment of ventilatory abnormality in the asymptomatic subject: an exercise in futility. *Thorax* 1966; 2: 445-449.
82. 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.
83. Miller MR, Pincock AC. Predicted values: how should we use them? *Thorax* 1988; 43: 265-267.
84. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948-968.
85. Miller MR, Quanjer PH, Swanney MP, Ruppel G, Enright PL. Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2011; 139: 52-59.
86. Perkins JM, Khan KT, Smith GD, Subramanian SV. Patterns and trends of adult height in India in 2005-2006. *Economics and Human Biology* 2011; 9: 184-193.
87. Gautam RK, Adak DK. Nutrition and Genetic Variation among Central Indian Tribes. In: Proceedings National Symposium of Tribal Health, 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>, last visited 24 March 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, Flores C, Cabrera VM. 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. Etler 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, Jin L, Case DT, Jiao YP, Wang XP, Bai CX, Jin G, Yang JM, Wang H, Yuan JB, Huang W, Wang ZG, Chen RB. Dermatoglyphics from all Chinese ethnic groups reveal geographic patterning. *PLoS One* 2010; 20; 5:e8783.
95. Crapo RO, Jensen RL, Oyunchimeg M, Tsh T, DuWayne Schmidt C. 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. *The China Journal* 2000; 44: 1-39.
97. Stature and Famine in China: The Welfare of the Survivors of the Great Leap Forward Famine, 1959-61, <http://ehsanz.econ.usyd.edu.au/papers/Morgan.pdf>. Last accessed 24 March 2012.
98. Morgan SL, 1998. 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*. Franz Steiner Verlag, pp. 7-34.
99. Lin W-S, Zhu F-C, Chen ACN, Xin W-H, Su Z, Li J-Y, Ye G-S. 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, Kumakura C, Zhou X, Jin F, Li Y, Lu S. Growth of height and leg length of children of Beijing and Xilinhot, China. *Anthropol Sci* 2008; 116: 76-76.
101. Weitz CA, Garruto RM, Chin C, Liu J. 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, Tse LY, Huen KF, Wong GWK, Law WY, Yeung VTF, Yeung WKY, Leung NK. 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. Growth charts from Korean Agency for Technology and Standards, Ministry of Knowledge Economy, South Korea. <http://sizekorea.kats.go.kr/> (Korean). Courtesy dr. Yeon-Mok Oh.
105. Swanney MP, Jensen RL, Crichton DA, Beckert LE, Cardno LA, Crapo RO. 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. <http://www.thoracic.org.au/accreditation/>), accessed 24 March 2012.
107. American Thoracic Society. Standardization of spirometry, 1994 update. *Am J Respir Crit Care Med* 1995; 152: 1107–1136.
108. Tanner, J.M., 1986. Growth as a target-seeking function: catch-up and -down growth in man. In: Falkner, F., Tanner, J.M. (Eds.), *Human Growth: A Comprehensive Treatise*, vol. 2. Plenum Press, New York, 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, Woolcock AJ. What factors explain racial differences in lung volumes? *Eur Respir J* 1991; 3: 829-838.

117. White NW, Hanley JH, Lalloo UG, Becklake MR. 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, Donnelly P, Unger W, Woolcock A. 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, Tang H. 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: 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, Stein MB, Shriver MD, Gelernter J. 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. Zapletal A, Paul T, Samanek N. Die Bedeutung heutiger Methoden der Lungenfunktionsdiagnostik zur Feststellung einer Obstruktion der Atemwege bei Kindern und Jugendlichen. *Z Erkrank Atm-Org* 1977; 149: 343-371.
132. Rosenthal M, Bain SH, Cramer D, Helms P, Denison D, Bush A, Warner JO. Lung function in white children aged 4-19 years: I - Spirometry. *Thorax* 1993; 48: 794-802.
133. Fulambarker A, Copur AS, Cohen ME, Patel M, Gill S, Schultz ST, Quanjer PH. Comparison of pulmonary function in immigrant vs US-born Asian Indians. *Chest* 2010; 137: 1398-1404.
134. ATS Statement. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Resp Dis* 1991; 144: 1202-1218.
135. Corey PN, Ashley MJ, Chan-Yeung M. Racial differences in lung function: search for proportional relationships. *J Occup Med* 1979; 21: 395-398.
136. http://www.lungfunction.org/enquiry_stanojevic_implementation_phq.pdf, last accessed 24 March 2012.
137. 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.
138. Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B: Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis* 1983; 127: 725-734.
139. Levy ML, Quanjer PH, Booker R, Cooper BG, Holmes S, Small IR. 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. DOI: <http://dx.doi.org/10.4104/pcrj.2009.00054>.

Table 1 – Summary of datasets included in the initial analysis. Total sample 97,759.

Group	Countries	Males		Females	
		N	Age range	N	Age range
African American	1	1,529	6-85	2,029	6.1-87
India + Pakistan	2	2,837	4-86	3,003	3-79
Latin America	5	2,337	6.7-89.4	2,578	7.4-89.7
Mexican American	1	1,622	6.2-86	2,282	6.5-87
Iran	1	3,398	5-85	2,739	5-80
Oman	1	638	6-65	618	6-65
North East Asia	2	2,176	15.3-91	4,526	15.5-90
South East Asia	4	4,187	3.3-88	6,371	3.1-92
North Africa	2	541	6-78	602	6-90
Caucasian	14	24,229	2.5-95	28,844	2.5-95
Other	-	199	6.2-93	474	5.8-91
Total	33	43,693	2.5-95	54,066	2.5-95

Table 2 – Number and age ranges of subjects in the groups for which prediction equations were derived. The countries of origin are listed in OLS Table E1.

Group	Males		Females		Total
	N	Age (yr)	N	Age (yr)	
Caucasian	25,827	2.5-95	31,568	2.5-95	57,395
African American	1,520	6-85	2,025	6.1-87	3,545
North East Asia	1,414	16-91	3,578	16-88	4,992
South East Asia	3,095	3.3-86	5,160	3.2-92	8,255
Total	31,856		42,331		74,187

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 Asia	-0.7	-2.1	1.1	-7.7	-2.7	-3.6	0.9	-3.2
South East Asia	-13.0	-15.7	2.9	-4.3	-9.7	-12.3	2.8	-0.9

Table 4 – Comparison of predicted values in adult Caucasian males according to different authors. The transition from one set of equations to another on the 18th birthday (a vs b or c vs d) is not smooth using older equations. The discontinuity, and the disparity between predicted values, is much worse in those short for age (a vs b), than in those of average height (c vs 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 years of age (e vs f).

a) 17.9 yr, 160 cm				b) 18.0 yr, 160 cm			
Author	FEV ₁	FVC	FEV ₁ /FVC	Author	FEV ₁	FVC	FEV ₁ /FVC
Polgar	3.09	3.42	N.A.	ECSC/ERS	3.67	4.23	0.83
Rosenthal	2.70	3.69	0.84	Hankinson	3.58	4.12	0.84
Zapletal	2.93	3.53	0.85	Stanojevic	3.46	4.09	0.85
Stanojevic	3.45	4.08	0.85	GLI 2012	3.61	4.13	0.88
Wang	3.49	3.93	0.87				
GLI 2012	3.61	4.12	0.88				

c) 17.9 yr, 180 cm				d) 18.0 yr, 180 cm			
Author	FEV ₁	FVC	FEV ₁ /FVC	Author	FEV ₁	FVC	FEV ₁ /FVC
Polgar	4.30	4.69	N.A.	ECSC/ERS	4.53	5.38	0.83
Rosenthal	4.27	5.17	0.82	Hankinson	4.53	5.39	0.84
Zapletal	4.11	4.99	0.85	Stanojevic	4.56	5.45	0.85
Stanojevic	4.55	5.44	0.85	GLI 2012	4.69	5.49	0.86
Wang	4.46	5.20	0.86				
GLI 2012	4.68	5.47	0.86				

e) 25 yr, 175 cm				f) 55 yr, 175 cm			
Author	FEV ₁	FVC	FEV ₁ /FVC	Author	FEV ₁	FVC	FEV ₁ /FVC
Crapo	4.45	5.32	0.84	Crapo	3.71	4.67	0.79
ECSC/ERS	4.31	5.09	0.83	ECSC/ERS	3.44	4.31	0.77
Hankinson	4.44	5.36	0.83	Hankinson	3.63	4.74	0.77
HSE	4.43	5.29	0.85	HSE	3.60	4.63	0.79
Knudson	4.39	5.24	0.84	Knudson	3.52	4.35	0.81
Stanojevic	4.42	5.36	0.83	Stanojevic	3.63	4.75	0.77
GLI 2012	4.46	5.32	0.85	GLI 2012	3.65	4.66	0.79

Footnote: GLI = Global Lungs Initiative; Crapo [137], ECSC/ERS [8]; Hankinson – NHANES III [17]; HSE = Health Survey for England 1995-1996 [53]; Knudson [138]; Polgar [7]; Rosenthal [132]; Stanojevic- 'all age' [13]; Zapletal [131], Wang [6]; N.A. = not available.

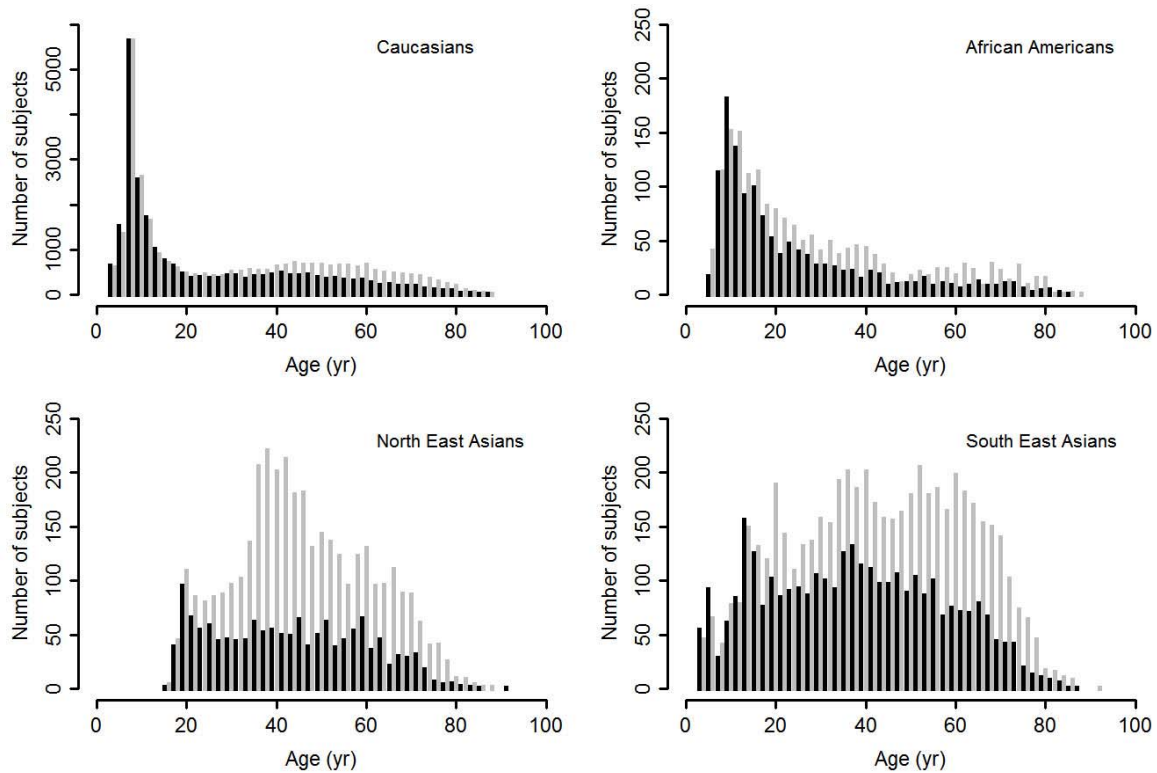


Figure 1 – Age distribution in the final sample (N=74,187). Note that the Y-axis in Caucasians differs from that in the other groups. The large number of 8-year olds did not lead to any bias in predicted values [14].

Black = males, grey = females.

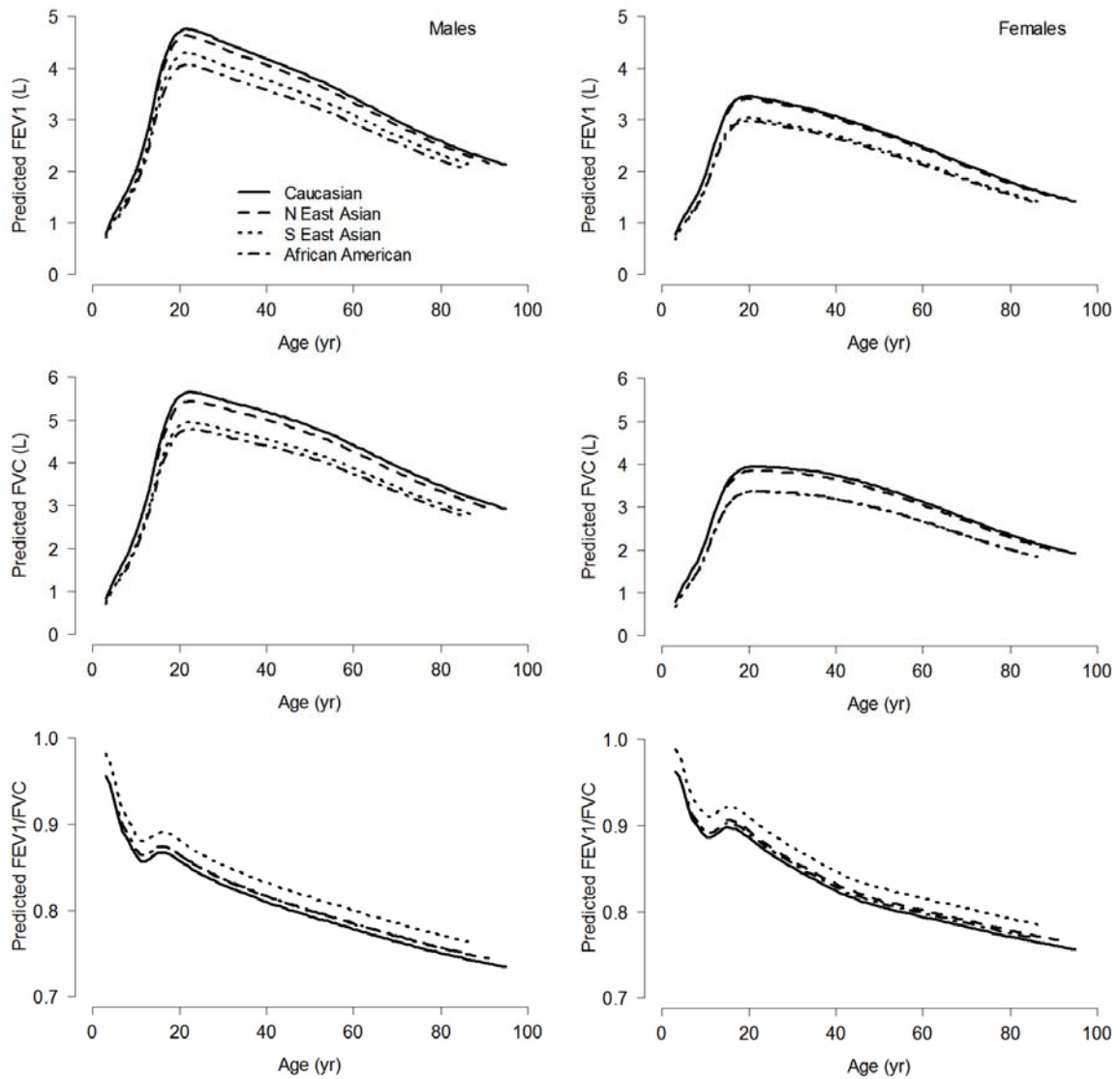


Figure 2 – Predicted values for FEV₁, FVC and FEV₁/FVC by sex and ethnic group. Note that in some cases lines overlap (*cf.* Table 3). Graphs were generated using mean height for age in Caucasians to illustrate proportional differences between ethnic groups of the same height and age; in practice differences in height for age further affect predicted values. The dip and rise in FEV₁/FVC around adolescence is due to differential changes in FEV₁ and FVC [15].

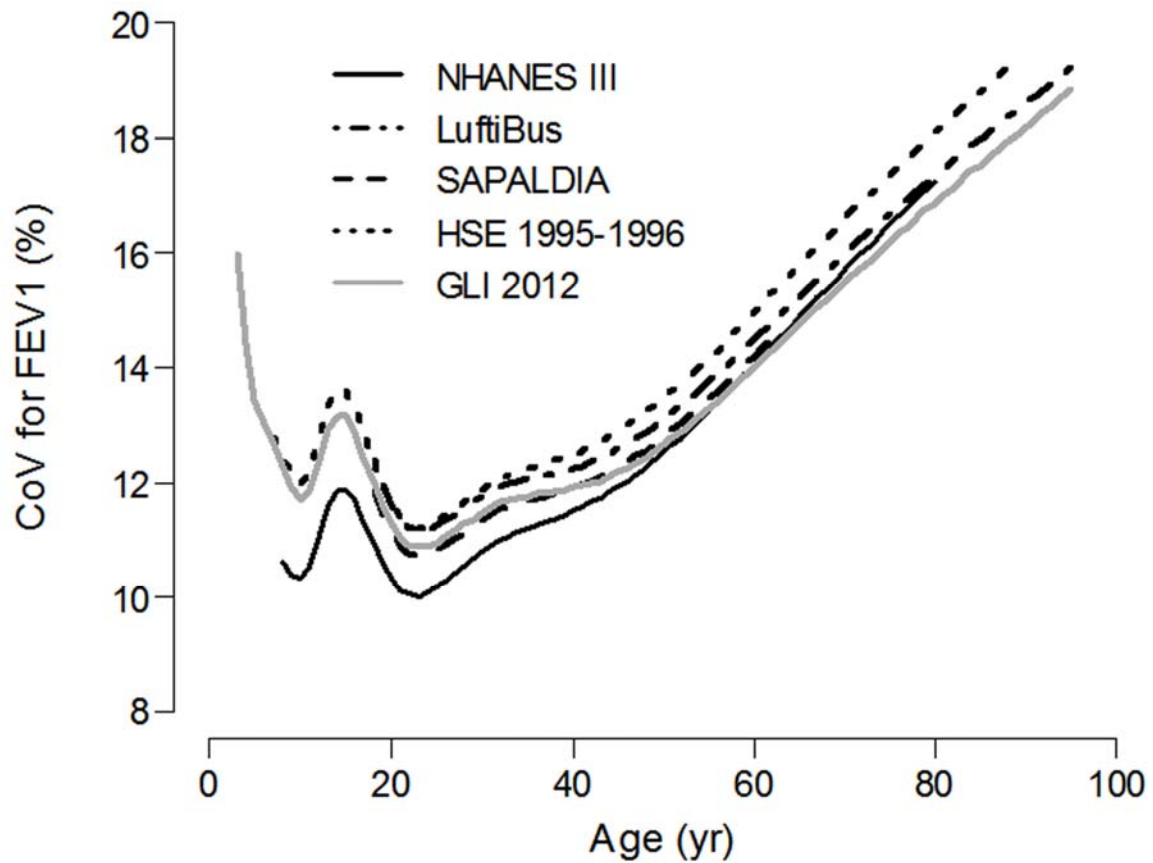


Figure 3 – Coefficient of variation (CoV) for FEV₁ in healthy Caucasian males in 4 large, recent studies, and in the collated GLI material, shown by the solid gray line. Graphs were generated using mean height for age in Caucasians to illustrate proportional differences between ethnic groups of the same height and age; in practice ethnic differences in height for age further affect predicted values. Findings for NHANES IV [31] and NHANES III were practically identical.

Footnote: NHANES III [17]; LuftiBus [54], SAPALDIA [52]; HSE 1995-1996 [53]; GLI 2012 = present study.

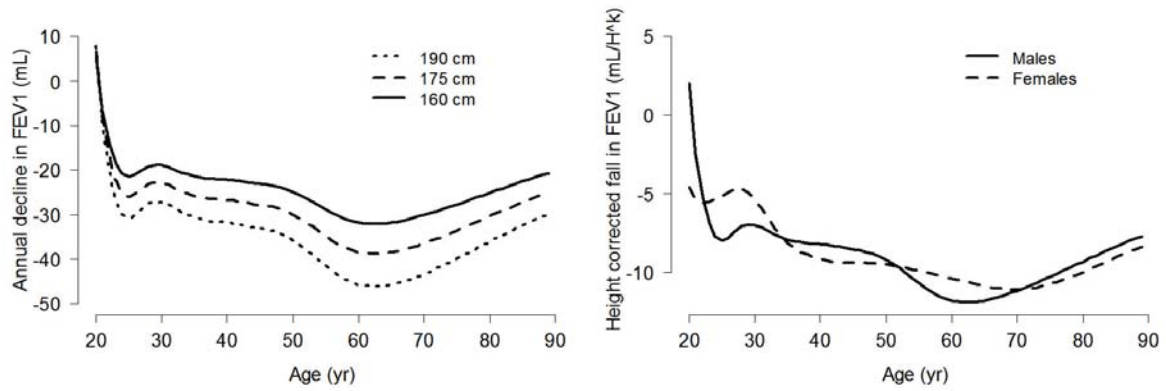


Figure 4 – The age-related cross-sectional annual change in FEV_1 is larger in tall than in short subjects; here shown for Caucasian males of height 160 cm, 175 cm and 190 cm (left panel). Standardising FEV_1 for height by taking the ratio $FEV_1/Height^k$, where $k \sim 2$, largely removes differences in age-related changes between adult males and females (right panel). Please note that these findings have no bearing on within-subject longitudinal decline.

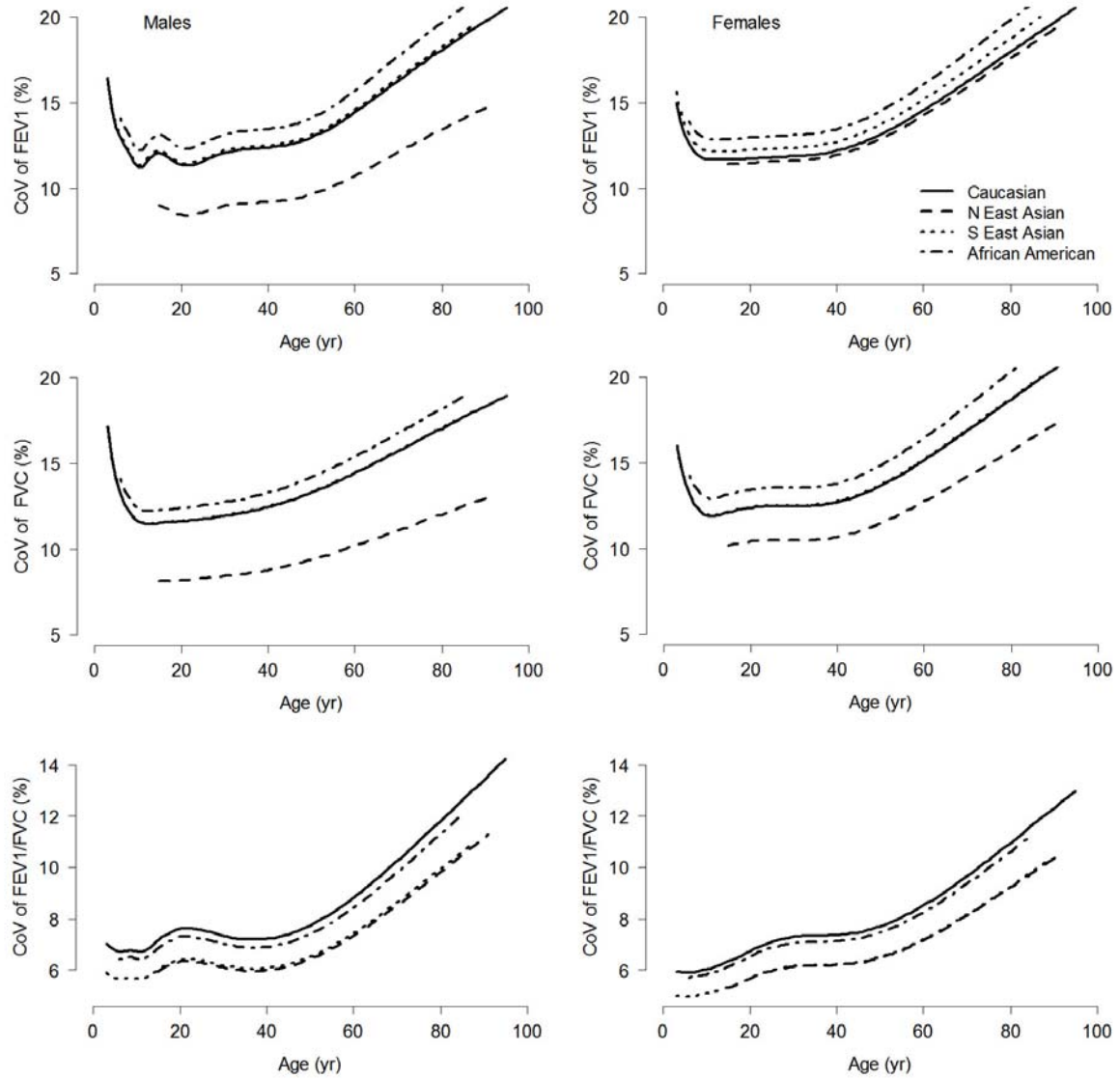


Figure 5 – Between-subject coefficients of variation (CoV) of FEV₁, FVC and FEV₁/FVC in Caucasians, African Americans, and North and South East Asians by age. Note that some lines overlap.

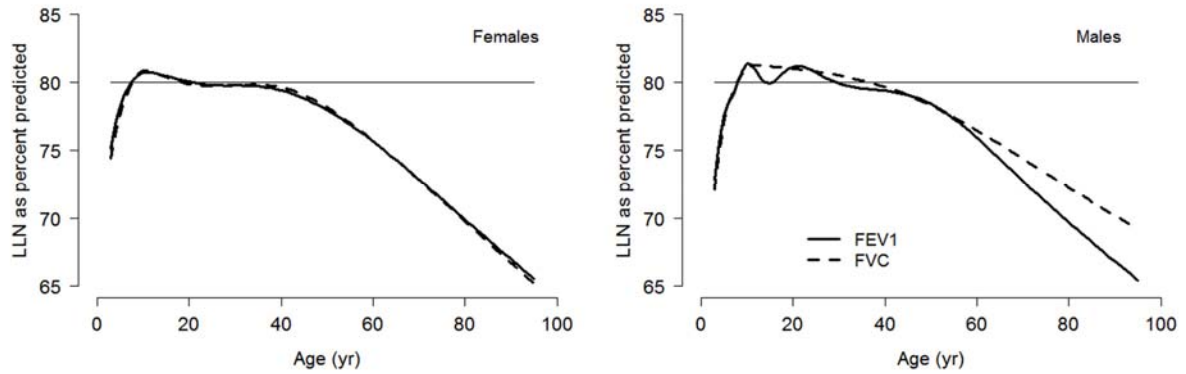


Figure 6 – Diagram illustrating the lower limit of normal (LLN), based on the 5th centile (-1.64 z-scores) values for FEV₁ and FVC, expressed as percent predicted, in Caucasian females and males. As can be seen, the use of 80% of predicted as the lower limit of normal gives rise to age-related bias. Graphs were generated using mean height for age in Caucasians.

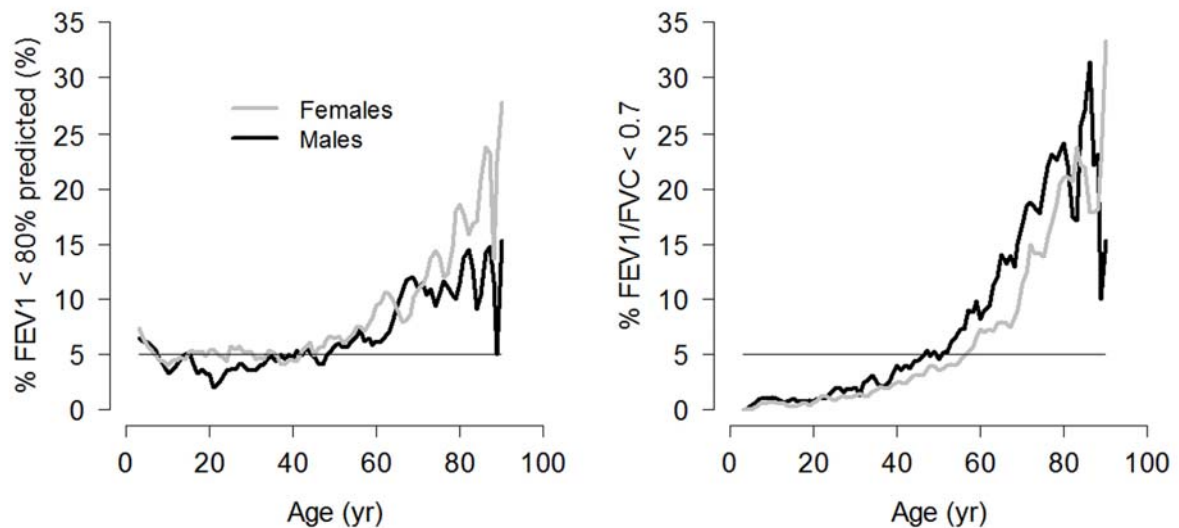


Figure 7 – Percentage of healthy male and female non-smokers (N=74,187) in whom FEV₁ is <80% predicted, or in whom FEV₁/FVC is less than 0.70. If it is assumed that FEV₁/FVC <0.70 (instead of below the 5th centile LLN) implies pathological airflow limitation, cases are missed in those under 50 years of age, whereas over diagnosis occurs above that age. Thus, at age 80 years between 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 (right panel).

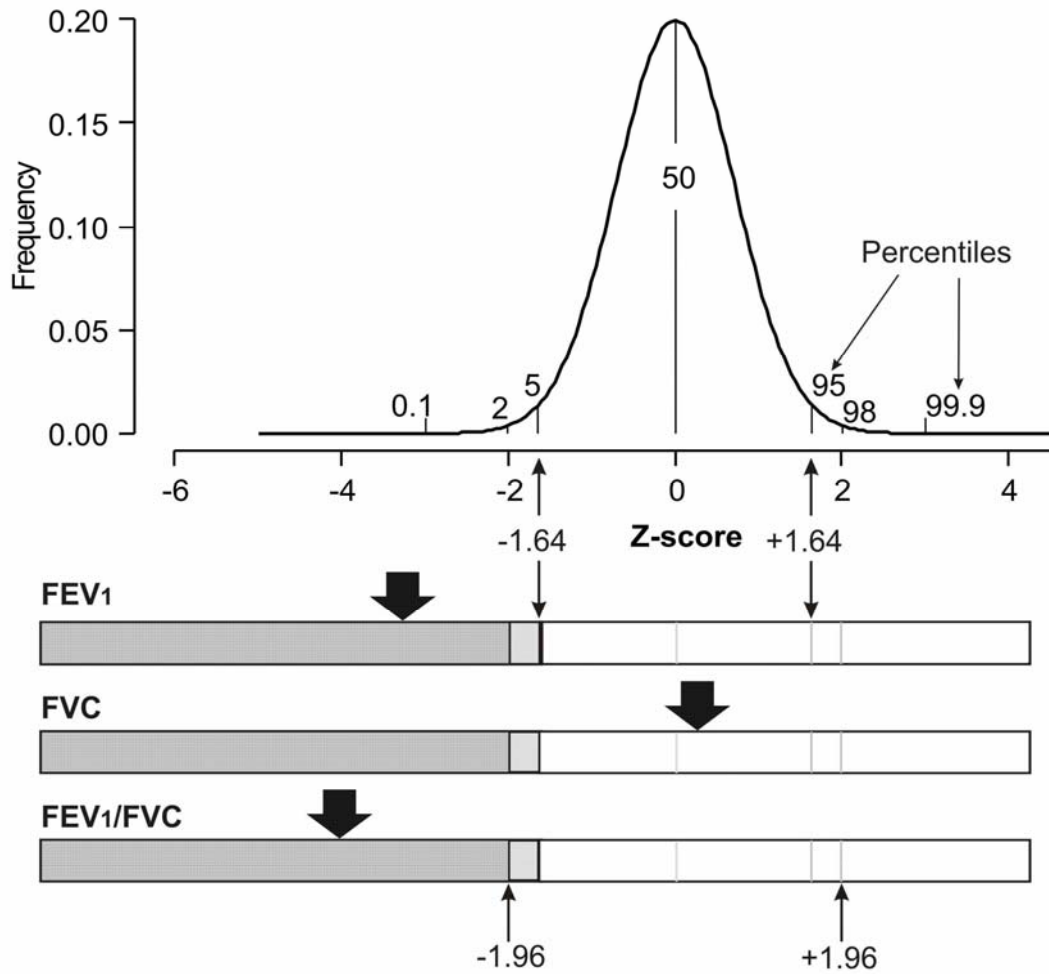


Figure 8 – Pictogram (bottom diagram) showing test results as z-scores (arrows) for a male from North China, age 34 years, height 166 cm.

	Measured	Predicted	LLN (5 th centile)	z-score
FEV ₁ (L)	2.56	3.66	3.03	-3.27
FVC (L)	4.47	4.39	3.78	0.22
FEV ₁ /FVC	0.57	0.84	0.75	-4.05

White area in the bars delineates observations that are above the mean -1.64 z-score (LLN 5%), light gray delimits mean -1.96 z-score (LLN 2.5%). The top diagram illustrates how z-scores relate to percentiles. Adapted from [139] with permission of the editors of the Primary Care Respiratory Journal.