

## **Influence of Secular Trends and Sample Size on Reference Equations for Lung Function Tests**

Philip H. Quanjer, Departments of Pulmonary Diseases and Pediatrics, Erasmus Medical Centre and Department of Paediatrics, Erasmus University, Rotterdam, the Netherlands

Janet Stocks, Portex Respiratory Unit, UCL Institute of Child Health, London, UK  
Tim J. Cole, MRC Centre of Epidemiology for Child Health, UCL Institute of Child Health, London, UK

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

Sanja Stanojevic, Portex Respiratory Unit, UCL Institute of Child Health, London, UK

on behalf of the ERS Global Lungs Task Force

### **Corresponding author:**

Philip Quanjer, Kervel 19, 7443 GT Nijverdal, the Netherlands. E-mail: [pquanjer@xs4all.nl](mailto:pquanjer@xs4all.nl)

Running title: **Adopting Reference Equations**

Title : **90 characters (90 allowed)**

**Word count manuscript: 3002 (3000 allowed)**

**Word count abstract: 200 (200 allowed)**

**Keywords:** multicentre study

**pulmonary function**

**reference equations**

**secular trends**

**spirometry**

***The ERS Global Lungs Initiative (see [www.lungfunction.org](http://www.lungfunction.org)):***

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

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

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

*Ip*

*Statistical reviewers:* M. Bottai, C. Schindler

*Persons and centres contributing data to this manuscript:* C.S. Beardsmore, Department of Infection, Immunity and Inflammation (Child Health), University of Leicester, Leicester, UK;

B. Brunekreef, Institute for Risk Assessment Sciences, Universiteit Utrecht, Utrecht, the Netherlands;

H. Eigen, Section of Pulmonology and Intensive Care, James Whitcomb Riley Hospital for Children, Indiana University School of Medicine, Indianapolis, IN, USA;

E. Falaschetti, [Health Survey for England 1995-1996 (HSE)], Department of Epidemiology and Public Health, UCL, London, UK;

B. Fallon, Respiratory Laboratory, Nepean Hospital, Penrith, Australia;

M.W. Gerbase, [and the SAPALDIA group, <http://www.sapaldia.net>], Division of Pulmonary Medicine, University Hospitals of Geneva, Geneva, Switzerland;

C.J. Gore, Physiology Department, Australian Institute of Sport, Belconnen, Australia;

G. 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;

M. Gappa, (and the LUNOKID study group) Children's Hospital and Research Institute, Marienhospital Wesel, Germany;

G.S. Kerby (Lung Function Measures in Preschool Children with Cystic Fibrosis study group), University of Colorado Denver School of Medicine Pulmonary Medicine, The Children's Hospital, Aurora, CO, USA;

J. Kühn, Klinik für Kinder- und Jugendmedizin, Städtisches Klinikum Karlsruhe, Karlsruhe, Germany;

S. Kuster, Lungenliga Zürich, Zürich, Switzerland;

A. Langhammer, [The HUNT Study] HUNT Research Centre, NTNU, Verdal, Norway;

S. Lum, Portex Respiratory Unit, UCL, Institute of Child Health, London, UK;

W. Nystad, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway;

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;

P.H. Quanjer, Departments of Pulmonary Diseases and Pediatrics, Erasmus Medical Centre – Sophia Children's Hospital, Erasmus University, Rotterdam, The Netherlands;

M. Rosenthal, Royal Brompton Hospital, London, UK;

S. Stanojevic, Portex Respiratory Unit, UCL, Institute of Child Health, London, UK (Asthma UK Growing Lungs Initiative (<http://www.growinglungs.org.uk>));

J.B. Soriano, Program of Epidemiology and Clinical Research, CIMERA, Recinte Hospital Joan March, Illes Balears, Spain (Framingham study);

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;

D. Vilozni, Pediatric Pulmonary Units of The Edmond and Lili Safra Children's Hospital, Sheba Medical Center Ramat-Gan and Meyer Children's Hospital of Haifa, Rambam Medical Center, Haifa, Israel;

H. Vlachos, Department of Pediatrics, Division of Respiratory Medicine, University of Sherbrooke, Quebec, Canada;

S. West, Respiratory Function Laboratory, Westmead Hospital, Australia;

D. Zagami, Lung Function Laboratory, Gold Coast Hospital, Southport, QLD, Australia.

## **Abstract**

**Aims:** Determine the contribution of secular trends and sample size to lung function reference equations, and establish the number of local subjects required to validate published reference values.

**Methods:** Thirty spirometry datasets, collected between 1978-2009, provided data on healthy whites: 19,291 males and 23,741 females, 2.5-95 years. The best fit for FEV<sub>1</sub>, FVC and FEV<sub>1</sub>/FVC as functions of age, height and sex were derived from the entire dataset using GAMLSS. Mean z-scores were calculated for individual datasets to determine inter-centre differences. This was repeated by subdividing one large dataset (3,683 males, 4,759 females) into 36 smaller subsets (comprising 18-227 individuals) to preclude differences due to population/technique.

**Results:** No secular trends were observed, and differences between datasets comprising >1000 subjects were small (maximum difference in FEV<sub>1</sub> and FVC from overall mean: +0.30 to -0.22 z-scores). Subdividing one large dataset into smaller subsets reproduced the above sample size related differences and revealed that at least 150 males and 150 females would be necessary to validate reference values to avoid spurious differences due to sampling error.

**Conclusions:** Use of local controls to validate reference equations will rarely be practical due to numbers required. Reference equations derived from large or collated datasets are recommended.

## INTRODUCTION

Tests of pulmonary function, in particular spirometry, play an important role in diagnosing obstructive lung disease [1-6], assessing the severity of lung disease, monitoring treatment of patients with respiratory disorders, and allocating patients to treatment groups in drug intervention studies. Since spirometric lung function depends on body size, age, sex and ethnic group, reference equations derived from healthy individuals are imperative for interpreting results. While there is ample choice of published equations [7], there are significant differences between reference equations, which have implications for the interpretation of results [8-14]. Furthermore, the observed differences between equations raise important questions about the causes of these discrepancies, how to select the most appropriate equations and whether or not 'local healthy controls' should be used to validate the selected reference.

It has been proposed [10] that the observed differences between prediction equations are due to technical and procedural differences (including quality control), true (biological) differences between populations, including secular trends, and chance differences between populations due to sampling ('sampling error'). Smaller samples have greater uncertainty about the true mean and thus increase the chance that they will differ from other datasets simply due to sampling error. Since many prediction equations are based on limited sample sizes, until now, it has been difficult to separate the effects of sampling error from those due to biological or technical differences.

Given the practical limitations of measuring large samples of healthy individuals to derive population specific reference equations, especially those derived from across the entire life span, it has been proposed that available datasets should be collated to produce globally applicable reference equations [13,15]. Earlier attempts to achieve this goal relied on constructing summary equations derived from published equations for specific age groups [16-18]. Although such summary equations have been widely used and served a useful role, their limitations soon became apparent [19]. The potential advantages of collating original data, and storing such data within an international database supported by the major international respiratory societies, was proposed as long ago as 1995 but has yet to be implemented [13]. The recent emergence of flexible and sophisticated statistical methods such as GAMLSS to model lung function data [20-22], provides a powerful tool to underpin such initiatives, and overcomes many of the previous difficulties, including those relating to the rapid age-related changes in childhood and adolescence when trying to derive 'all-age' equations [13,23-26]. The benefits of collating data include a larger sample size over a wider age range which will be more generalisable. This may, however, come at a cost of increased variability from outlying datasets due to the biological or technical differences mentioned above, which may in turn inflate the lower limit of normal (LLN). Furthermore, if secular trends in pulmonary function arise from a trend in physical development [27-28] and/or in technical improvements, equations based on collated datasets could be internally biased [29-34].

Although many users remain unaware of which prediction equations are being used to interpret their data, it is generally recommended to apply reference equations that have been derived from a similar population to that being tested, using comparable instruments, methods and quality standards [8,10,35]. It has also been suggested that results from a group of local healthy subjects should be used to validate the selected reference equations. However, in practice these recommendations are difficult to apply, since evidence regarding

the number of healthy subjects actually required is limited and inconsistent, estimates ranging from 20-100 [8-10,36].

The objectives of this study were to: (1) explore the differences between centres in a collated dataset, including those due to secular trends, and how these affect predicted values of spirometric lung function, (2) investigate the influence of sample size *per se* on the interpretation of lung function results, and (3) to estimate the minimum number of subjects required for local validation of reference equations.

## MATERIAL AND METHODS

Within the framework of the ERS Global Lungs Initiative [37] 64 centres from 28 countries across 5 continents have now shared spirometric data on 149,759 healthy, lifelong nonsmokers, aged 2.5-95 years. The overall aim of this group is to derive prediction equations for different ethnic groups valid worldwide. The analysis presented in this paper is confined to white subjects, as they currently represent the great majority of available data. Thirty centres provided data on 19,291 white males and 23,741 white females, age 2.5-95 years. The distribution by age is shown in Figure E1. Sample size in each of the individual datasets varied between 20-4759 subjects in females, and 17-3683 in males; 12 datasets comprised <100 females, and 10 had <100 males. Data were collected between 1978 and 2009. Although published data from the Health Survey for England 1995-1996 (HSE) study was limited to individuals 16 years and older [38], for the purposes of this analysis, these were complemented with data from 7-16 year olds that had been collected in the same study using identical protocol and techniques.

All datasets complied with international recommendations with regard to equipment, methods, procedures, data selection and quality control, valid at the time of data collection. All persons, groups or organisations sharing data with the Global Lungs Initiative have specified that data were collected with appropriate ethical approval, and permission has been given to publish results from the collated dataset.

GAMLSS [20-22] was used to derive the best fitting function for FEV<sub>1</sub>, FVC and FEV<sub>1</sub>/FVC as a function of age and height in males and females. The statistical methods used were as described by Cole *et al.* [23], details of which are provided in the online supplement (OLS). Once the equations had been derived from the entire dataset, they were used to calculate the average standardised residuals (z-scores) for each centre. Ideally, the residuals from each centre should have a mean of 0, and a standard deviation of one. To investigate the influence of secular trends, we displayed the residuals (z-scores) as a function of the year of measurement; in studies conducted over several years the midpoint of the study was used.

In order to explore the influence of sample size *per se* on predicted values, without any confounding effects due to biological or technical differences, and to estimate the minimal number of local controls that would be needed to validate published reference equations, we repeated this exercise after dividing data from the HSE study into smaller subsets. A random selection was performed (without replacement) so that 36 subgroups of males (N=3,683), subgroup size 38-225, and 46 subgroups of females (N=4,759), subgroup size 18-227 were formed. Using all the data in the HSE study, GAMLSS was again used to derive the best fitting function for FEV<sub>1</sub>, FVC and FEV<sub>1</sub>/FVC as a function of age and height, in women and men. Subsequently, the average standardised residuals (z-scores) were obtained for each of the 82 individual HSE subsets.

## RESULTS

### **Difference in mean z-score between centres**

Differences between the largest datasets were small and clinically minimal (Figure 1). The mean of the z-scores in the largest datasets was virtually zero. Maximum differences from the overall mean for FEV<sub>1</sub>, FVC and FEV<sub>1</sub>/FVC in datasets with >1000 subjects ranged from +0.30 to -0.22 z-score ( *i.e.* were within around 3% of that predicted from the collated dataset). For datasets with N>150 but <1000 the corresponding range for FEV<sub>1</sub> and FVC was +0.4 to -0.3 z-scores, with slightly larger differences for FEV<sub>1</sub>/FVC. The smaller the number of subjects in a dataset, the larger the offset from zero (Figure 1). Details of the numbers of subjects in each study, and the centre specific z-scores, are given in Table E1.

### **Secular trends**

There was no significant correlation between year of measurement and either sample size or the residuals for FEV<sub>1</sub>, FVC and FEV<sub>1</sub>/FVC (Figure 2, p=0.833 for FEV<sub>1</sub> and p=0.447 for FVC). Height differed between datasets from that in the HSE study by up to 2%; these differences were unrelated to the mean z-score for that centre.

### **Differences in scatter between centres**

The scatter (SD of the z-scores) in each centre was also related to sample size (Figure E2, OLS). Thus while the SD was fairly close to the ideal of 1 in datasets comprising >1000 subjects, there was increasing variability for sample sizes <1000, the extremes being 0.60 and 1.29. The average SD of the z-scores from all the centres was slightly less than 1 (0.97 in females, 0.95 in males for FEV<sub>1</sub>; 0.96 in females, 0.94 in males for FVC). The lower limit of normal in collated data is determined by the largest datasets; these have a slightly lower LLN for FEV<sub>1</sub> and FVC than in small datasets. This effect was more pronounced for the FEV<sub>1</sub>/FVC ratio (0.90 in both females and males). Further details are given in the OLS.

### **The effect of sample size on predicted values**

To investigate whether including the smaller datasets biased the predicted values and their scatter, we derived prediction equations that were limited to datasets where N>100 for each sex, and compared these to the values derived when data from all 30 centres had been included. This had minimal effect on either the predicted values (maximum difference in FEV<sub>1</sub> or FVC: 17 mL) or the scatter (maximum difference in SD up to 2 mL).

### **How big a sample size is needed to validate equations?**

To confirm whether differences observed between centres (Figure 1) were primarily due to sample size, we examined the differences within subsets of varying sizes from the HSE study. As can be seen from Figure 3, exactly the same pattern was observed as when examining differences in residuals *between* centres. Thus, there was minimal offset from the predicted value derived from the entire HSE dataset in the largest subsets (N > ~150), with increasing scatter in the residuals with decreasing group size (Figure 3). This is in keeping with results from other studies [39-41]. As these data derive from the same population and were collected by the same staff, using the same methods and quality control, differences can only be explained by 'sampling error'; the smaller the sample, the more the offset and variability from the expected mean.

The standard deviation of the z-scores was also calculated for each of the 82 subsets of data from the HSE study. The SD for FEV<sub>1</sub>, FVC and FEV<sub>1</sub>/FVC ranged between 0.80-1.21 in females, and 0.72-1.25 in males for the various subsets, the average being 0.99-1.00. Once

again, the pattern was similar to that observed from the 30 centres, *i.e.* the SD decreased as sample size increased, with close agreement once the sample size was >150. This is in keeping with results from other studies [39-41]. These findings suggest that at least 300 local healthy controls (150 males and 150 females) would be needed to validate published reference equations with any degree of certainty, since with smaller sample sizes differences of up to 0.5 z-scores may occur purely by chance.

## DISCUSSION

This study provides evidence with respect to several important issues related to the interpretation of spirometry data using reference equations. Firstly, differences between large datasets were clinically insignificant, and no evidence was found of secular trends in data collected from white subjects over the past 30 years. Secondly, reference equations derived from collated datasets are a reasonable approach to interpret lung function results across different centres. Thirdly, differences between datasets based on a relatively homogeneous population (in this case white subjects of European origin) are largely explained by sampling variability. Fourthly, validation of published reference values for spirometry using local controls is unlikely to be practical due to the numbers required to avoid sampling error. Finally, prediction equations derived from less than 150 individuals for each sex are unlikely to be reliable. This is likely to hold true for any ethnic group.

Despite being collected over a period of 30 years (1978-2009), there was no evidence of a secular trend in pulmonary function (Figure 2). This strongly suggests that differences between centres can only be attributed to a minor extent to biological and technical differences. Secular trends in pulmonary function have been described previously [28-34]. They may be due to a cohort effect, *i.e.* that persons born for example 50 years ago had a different lung development during growth than those born 30 years ago. Infectious diseases, nutrition, smoking during pregnancy, exposure to environmental factors, different age of maturation, *etc.* might all leave their marks so that the lung function of a 50 year old person born in 1930 would be different from that in a 50 yr old person born in 1960. The fact that we did not find evidence for a secular trend trend in spirometric indices over the past 25-30 years may be due to the fact that the present study is based on data derived from societies in Europe, North America, Australia and New Zealand with comparable and relatively stable socio-economic conditions. Alternatively, since secular changes in height with increasing affluence tend to reflect changes in leg length rather than trunk height [27], it is possible that subtle improvements in lung function over time may be masked when lung function is based on standing height. Since data in this study complied with different standards issued by ECCS, ECCS/ERS, ATS and ATS/ERS [17,35,42-44], secular trends might spuriously arise from the application of more rigorous standards and differences in instruments used. For example, Hankinson *et al.* [45] found that applying the 1994 ATS [43] rather than the 1987 standards [42] made a difference in the level of FEV<sub>1</sub> and FVC equivalent to z-scores of ~0.14 and ~0.12, respectively. Our results suggest that any trend due to implementation of more rigorous standards is obscured by small differences between populations. This may not hold true for measurements of lung volumes and T<sub>L,CO</sub>.

A major concern when working with collated data is that differences in inclusion criteria or in methodological and other issues will adversely affect the accuracy and precision of predicted values, *i.e.* the average and the lower limits of normal (LLN). Our findings suggest that biological differences between populations were small, that standards of administering pulmonary function tests and quality control were relatively uniform and that differences between centres seem to be mainly due to sampling error. By creating subgroups of different sizes from the large HSE study, we controlled for secular trends, biological differences and

technical differences in administering pulmonary function tests, as well as in quality control. This analysis confirmed that the offset from predicted values is related to sampling error, such that smaller sample sizes have greater offset from the predicted mean. This provides further evidence to support the collation of data in order to derive more robust prediction equations.

When combining data from 30 sources around the world, we would expect greater offset and variability than in the sub-samples of datasets derived from a large population (HSE). Instead, there was remarkable similarity, with only the smallest datasets differing more from the overall mean, than the sub-samples from the HSE population (Figure 3). However, the findings differed with respect to the residual scatter, which was on average smallest in the smallest datasets (Figure E2), indicating that populations in the largest datasets were less homogeneous. Because the results of regression analysis are dominated by the largest datasets, which reflect a representative population sample, the LLN derived from the collated data will be slightly lower than those in datasets comprising < 1000 individuals (Figure E2). As the largest datasets arose from random population samples, the predicted values and their LLNs are representative of such populations. Our results also indicate that while inclusion of the smaller datasets does not lead to biased results, predicted values and LLNs derived from small local datasets of healthy controls are unlikely to be representative of a general population. This does not diminish the value of recruiting prospective controls in research studies.

In the past, equations based on collated data were generated using published prediction equations [16-18,35]. This leaves much to be desired as the prediction equations were simple in structure and therefore did not allow appropriate modelling of the complex relationship of pulmonary function with age and height from derived data. This also holds for determining the LLN. A clear advantage of deriving prediction equations from collated data using modern statistical techniques and the actual results from each individual is that the age-related changes in the level of pulmonary function as well as the LLN can be modelled properly [15,23,25-26].

Our findings have practical implications, in that laboratories seeking to validate reference equations for spirometry need to collect data on at least 300 healthy subjects. Even then, results from a local reference population are still likely to differ from predicted values by up to 0.2 z-scores. Since the between-subject variability (coefficient of variation) in this study varied from about 11% in adolescents to 17% in elderly subjects and those under 6 years of age, a difference of 0.2 z-scores equate to 2-3.5% of predicted, depending on the age of the local reference group. Thus, while validation of reference equations is potentially important, it appears to be largely impractical. Given the cost and effort required to form such large local reference populations, it is probably more practical and acceptable for laboratories to adhere closely to international standards for pulmonary function testing and adopt reference equations derived from large or collated studies.

## CONCLUSIONS

The collation of spirometric data to produce reference equations is a robust approach to interpret lung function results. The potential and minimal inflation of the limits of normality when using a collated dataset are balanced against a greater bias in predicted values when smaller datasets are used. Differences between centres are unlikely to be due to secular trends, technical or population differences, but rather sampling variability due to small sample sizes. The validation of external reference values by lung function laboratories is



quite costly and impractical. In practice it is advisable for laboratories which adhere closely to international standards for pulmonary function testing to adopt reference equations derived from large studies. This study provides evidence in support of the collation of spirometry data to create world-wide all- age reference equations.

## **Acknowledgements**

The Global Lungs Initiative is a European Respiratory Society Task Force.

Graham Hall was funded by Princess Margaret Hospital Foundation.

Sanja Stanojevic was funded by Asthma UK and in part by the Klosterfrau Foundation.

The following individuals/centres share datasets (occasionally more than 1) with the Global Lungs Initiative, centres that formed the Asthma UK initiative are marked with an asterisk:

O. Al-Rawas, Oman; M. Badier, France\*; C. Beardsmore, UK\*; H. Ben Saad, Tunisia; B. Brunekreef, the Netherlands; P. Burney, UK; W. Dejsomritrutai, Thailand; H. Eigen, USA\*; B. Fallon, Australia; A.M. Fulambarker, USA; M. Gappa, Germany\*; M. Gerbase, Switzerland; M. Golshan, Iran; C. Gore, Australia; G. Hall, Australia\*; J.L. Hankinson, USA\*; J. Henderson, UK; M.S.M. Ip, China; S. Karrasch, Germany; G. Kerby, USA\*; J. Kühn, Germany; S. Kuster, Switzerland; A. Langhammer, Norway; S. Lum, UK\*; A.L. Miller, USA; W. Nystad, Norway\*; Y.M. Oh, South Korea; W-H. Pan, Institute Taiwan ROC; R. Perez-Padilla, Mexico; P. Piccioni, Italy\*; F. Pistelli, Italy; Prasad KVV, India; P.H. Quanjer, the Netherlands; M. Rosenthal, UK\*; S. Stanojevic, UK\*; J.B. Soriano, Spain; F. Thomas, France; W. Tomalak, Poland\*; Y. Trabelsi, Tunisia; S.W. Turner, UK\*; D. Viložni, Israel\*; H. Vlachos, Canada\*; R. Warshaw, USA; S. West, Australia; D. Zagami, Australia; J.P. Zheng, China\*.

## References

1. Pellegrino R, Viegi G, Brusasco V *et al.* Interpretative strategies for lung function tests. Series “ATS/ERS Task Force: Standardisation of lung function testing”. *Eur Respir J* 2005; 26: 948-968.
2. Global strategy for the diagnosis, management and prevention of chronic obstructive lung disease. In GOLD\_Report.pdf on [www.goldcopd.com](http://www.goldcopd.com), accessed Jan. 4, 2010.
3. ATS/ERS Task Force: Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004; 23: 932–946.
4. BTS Guidelines for the Management of chronic obstructive pulmonary disease. *Thorax* 1997; 52: 1-28.
5. National Institute for Clinical Excellence. Clinical Guideline 12. Chronic obstructive pulmonary disease, February 2004. [www.nice.org.uk/CG012NICEguideline](http://www.nice.org.uk/CG012NICEguideline).
6. Global Strategy for Asthma Management and Prevention, Global Initiative for Asthma (GINA) 2008. Available from: <http://www.ginasthma.org>. Accessed July 10, 2010.
7. <http://www.spirxpert.com/GOLD.html>, downloading the software gives access to over 80 sets of prediction equations and an extensive help file with regression equations and some details of the studies.
8. ATS Statement. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Resp Dis* 1991; 144: 1202–1218.
9. ATS/ERS Statement. Pulmonary function testing in preschool children. *Am J Respir Crit Care Med* 2007; 175: 1304–1345.
10. ATS/ERS Task Force: Standardisation of lung function testing. Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26, 948-968.
11. Swanney MP, Ruppel G, Enright PL *et al.* Using the lower limit of normal for the FEV<sub>1</sub>/FVC ratio reduces the misclassification of airway obstruction. *Thorax* 2008; 63; 1046-1051.
12. Miller MR, Quanjer PH, Swanney MP, Ruppel G, Enright PL. Interpreting lung function data using 80 percent of predicted and 5 fixed thresholds misclassifies over 20% of patients. *Chest* 2010: in press.
13. Quanjer PH, Borsboom GJJM, Brunekreef B *et al.* Spirometric reference values for white European children and adolescents: Polgar revisited. *Pediatr Pulm* 1995; 19: 135-142.
14. Quanjer PH, Borsboom GJJM, Kivastik J, *et al.* Cross-sectional and longitudinal spirometry in children and adolescents. Interpretative strategies. *Am J Respir Crit Care Med* 2008; 178: 1262–1270.
15. 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.
16. Polgar G, Promadhat V. Pulmonary Function Testing in Children: Techniques and Standards. Philadelphia, Saunders, 1971.
17. Quanjer PhH (ed.) Standardized lung function testing. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. *Bull Eur Physiopathol Respir* 1983; 19 suppl. 5: 45-51.
18. Stocks J, Quanjer PH. Reference values for residual volume, functional residual capacity and total lung capacity. ATS Workshop on lung volume measurements. Official statement of the European Respiratory Society. *Eur Respir J* 1995; 8: 492–506.
19. Baur X, Isringhausen-Bley S, Degens P. Comparison of lung function reference values. *Int Arch Occup Environ Health* 1999; 72: 69-83.

20. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with Discussion). *Applied Statistics* 2005; 54: 507–544.
21. Mikis Stasinopoulos and Bob Rigby with contributions from Calliope Akantziliotou. (2008). GAMLSS: Generalized Additive Models for Location Scale and Shape. R package version 1.9-4. <http://www.gamlss.com/>.
22. R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
23. 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.
24. Stanojevic S, Wade A, Stocks J. Reference values for lung function: past, present and future. *Eur Respir J* 2010; 36: 12-19.
25. Stanojevic S, Wade A, Cole TJ, *et al.* on behalf of the Asthma UK collaborative group. Spirometry centile charts for young Caucasian children: The Asthma UK Collaborative Initiative. *Am J Respir Crit Care Med*, 2009; 180: 547-552.
26. Quanjer PH, Stanojevic S, Stocks J, *et al.* Changes in the FEV<sub>1</sub>/FVC ratio during childhood and adolescence: an intercontinental study. *Eur Respir J*, published online March 29, 2010 as doi: 10.1183/09031936.00164109.
27. Tanner JM, Hayashi T, Preece MA, Cameron N. 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.
28. Ip MS, Karlberg EM, Karlberg JP, *et al.* Lung function reference values in Chinese children and adolescents in Hong Kong: I. Spirometric values and comparison with other populations. *Am J Respir Crit Care Med* 2000; 162:424–429
29. 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.
30. Burrows B, Lebowitz MD, Casmilli 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.
31. Ware JH, Dockery DW, Louis TA, Xu XP, Ferris BG Jr, Speizer FE. Longitudinal and cross-sectional estimates of pulmonary function decline in never-smoking adults. *Am J Epidemiol* 1990; 132: 685-700.
32. Van Pelt W, Borsboom GJJM, Rijcken B, Schouten JP, van Zomeren BC, Quanjer PhH. 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.
33. 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.
34. Ip MS, Karlberg EM, Chan KN, *et al.* Lung function reference values in Chinese children and adolescents in Hong Kong: II. Prediction equations for plethysmographic lung volumes. *Am J Respir Crit Care Med* 2000; 162:430–435.
35. 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.
36. Jensen RL, Crapo RO, Flint AK, Howell HM. Problems in selecting representative reference values for spirometry. *Am J Respir Crit Care Med* 2002; 165: A200.
  37. Information about the Global Lungs Initiative is available from [www.lungfunction.org](http://www.lungfunction.org).
  38. 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.
  39. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol* 2003; 56: 441-477.
  40. Peek N, Arts D, Bosman R, van der Voort P, de Keizer N. External validation of prognostic models for critically ill patients required substantial sample sizes. *J Clin Epidemiol* 2007; 60: 491-501.
  41. Knofczynski GT, Mundfrom D. Sample sizes when using multiple linear regression for prediction. *Educational and Psychological Measurement* 2008; 68: 431-442.
  42. American Thoracic Society. Standardization of spirometry: 1987 update. ATS Statement. *Am Rev Respir Dis* 1987; 136: 1285–1298.
  43. American Thoracic Society. Standardization of spirometry: 1994 Update. ATS Statement. *Am J Respir Crit Care Med* 1995; 152: 1107–1136.
  44. Miller MR, Hankinson J, Brusaco V, *et al.* ATS/ERS statement. Standardization of spirometry. *Eur Respir J* 2005; 26: 319–338.
  45. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1995; 152: 179-187.

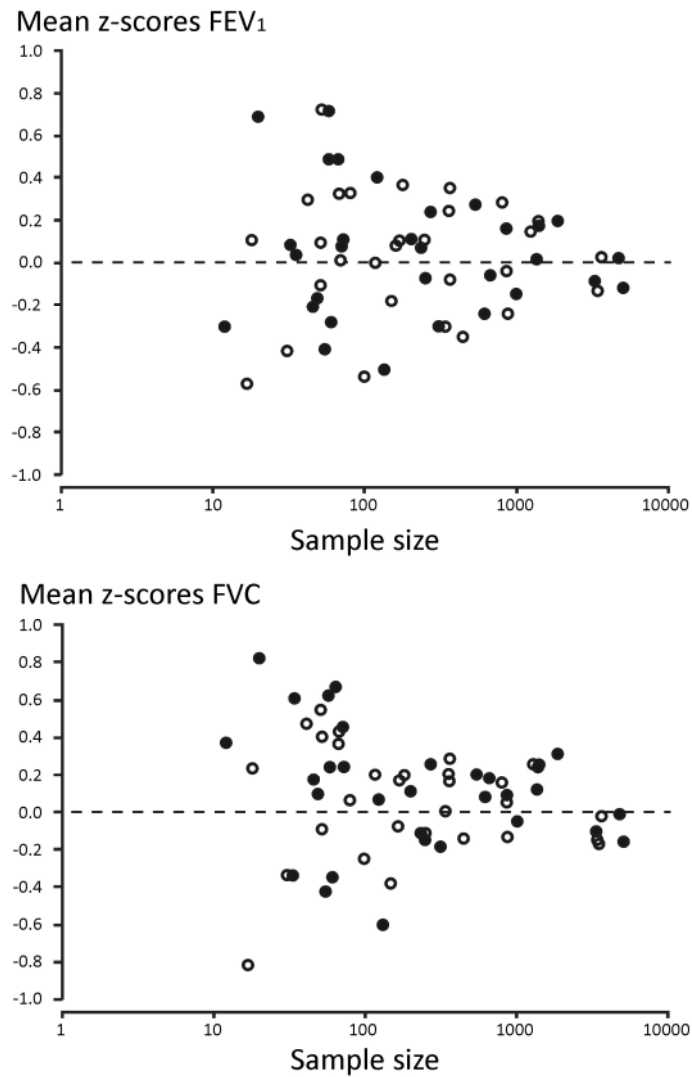


Figure 1 – Distribution of mean z-scores (standardised residuals) in 30 datasets, as a function of sample size. Females open circles, males closed circles.

Legend/footnote: z-scores for each centre were calculated using the equations that were derived from the entire collated dataset by using GAMLSS to regress FEV<sub>1</sub> and FVC on age and height, separately for males and females.

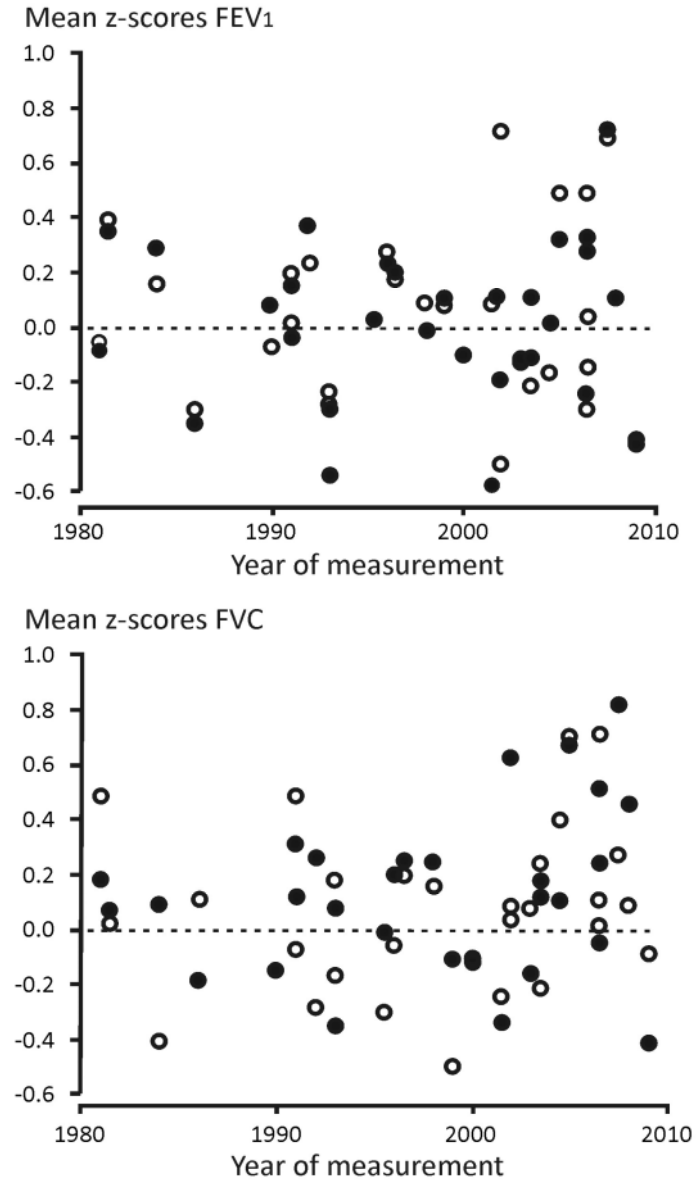


Figure 2 – Distribution of mean z-scores for FEV<sub>1</sub> and FVC in 30 datasets, as a function of the year of measurement. Females open circles, males closed circles.

Legend/footnote: As can be seen, no secular trends were evident in data collected from white subjects over the past 30 years.

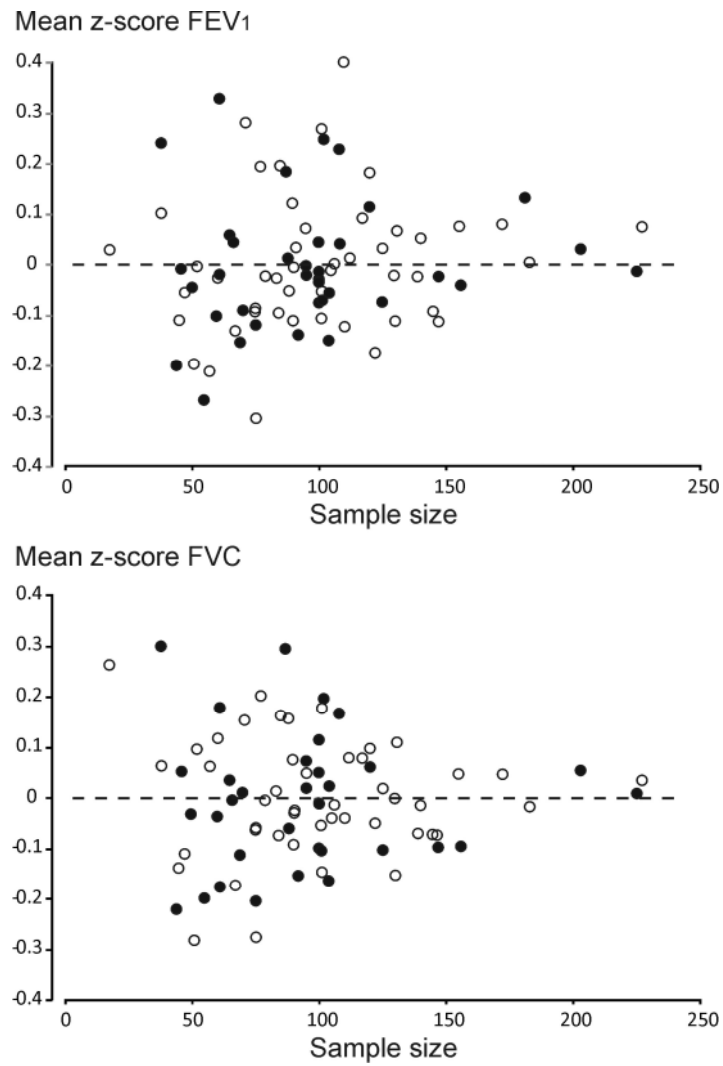


Figure 3 – Distribution of mean z-scores (standardised residuals) in 3,683 males and 4,759 females, as a function of sample size in subsets of the HSE [38] population. Females open circles, males closed circles.