GLI-2012
All-Age Multi-Ethnic
Reference Values for Spirometry

Advantages
Consequences

Philip H. Quanjer
Sanja Stanojevic
Janet Stocks
Tim J. Cole

Interpretation of spirometric data

Philip H. Quanjer
Sanja Stanojevic
Janet Stocks
Tim J. Cole

Introduction

In the four years that it took the Global Lung Function Initiative (GLI) to finish its mission, with the support of six large international respiratory societies, a collaborative network was established that spanned the world. The network included clinicians, researchers, technicians, IT engineers and manufacturers. The objective was to derive reference equations for spirometry that covered as many ethnic groups as possible, and an age range from pre-school children to old age. Thanks to unprecedented international cooperation tens of thousands records of spirometric measurements from healthy, non-smoking males and females, were made available by some 70 centres and organisations. These data were collated and analysed with modern statistical techniques, and led to the GLI-2012 prediction equations. This manuscript summarises the main results that have been previously presented at international meetings and in print.

Historical perspective

It took a long time before the introduction of the use of the spirometer by Hutchinson in 1846 [1] led to clinical applications. Inasmuch as it was clinically applied, measurements were limited to the assessment of "vital" capacity (VC), *i.e.* the slow expiratory vital capacity (EVC) according to today's terminology. Figure 1 illustrates the subdivision of the total lung capacity in EVC and residual volume in Hutchinson's publication. It took one century before the French investigators Tiffeneau and Pinelli [2] transformed spirometric measurements to the present form, in which the forced expiratory volume in 1 second (FEV1) and the inspiratory or forced expiratory VC (IVC and FVC) became pivotal diagnostic indices in clinical medicine. Yernault summarised the history of spirometric measurements concisely in a clear and accessible publication [3].

Spirometric test results are significantly influenced by subject cooperation, and are affected by technical factors; it follows that measurements need to be administered according to a strict protocol. In 1960 the European Community for Coal and Steel (ECCS) was the first organisation to issue recommendations [4]. This was followed by an update in 1971 [5], which comprised predicted values for spirometric indices, residual volume, total lung capacity and functional residual capacity. A few years later the first efforts at standardisation were made in the United States, initially only
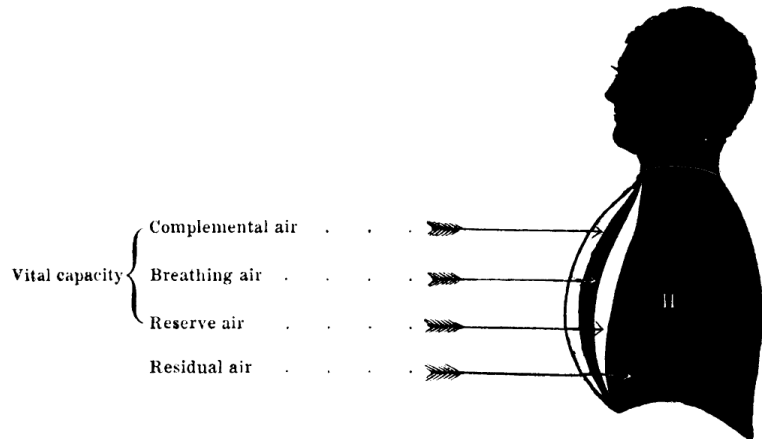

DIAGRAM 1.
The division of the thoracic movements.

Fig. 1 - Subdivision of the total lung capacity according to Hutchinson (1846).

for spirometry in an epidemiological setting [6-7]. Due to rapid technological developments, increased insight in the pathophysiology of lung diseases, and a greater arsenal of clinical lung function tests, a revision of the ECCS report was soon called for [8]. From then on revised standardisation reports were issued in the United States and Europe; American reports dealt with spirometry only, European recommendations covered a wider range of lung function tests and were invariably combined with recommended sets of reference values [9-11].

Reference values

The sets of reference values issued by the ECCS [4-5] were based on males working in coal mines and steel works. This was not a representative reference population, and in practice the predicted values were deemed to be too high. Even though no women had been tested, the ECCS issued reference values for females: they were 80% of the values for males. In 1983 the ECCS declined allocating funds for a population study to derive reference values obtained with methods that complied with the latest standards. With a view to combining technical recommendations with appropriate prediction equations, and because no material was available that had been obtained with appropriate techniques, for lack of better alternatives the standardisation committee decided to adopt the technique previously applied by Polgar [12] when deriving reference equations for children. This entailed the generation of a set of predicted values for age, height and sex using published prediction equations, and using this artificially generated set to derive new regression equations. Serious objections can be raised

against this procedure, but the resulting regression equations were accepted with scarcely any criticism and subsequently widely adopted.

An alternative that the ECCS standardisation group would have welcomed as a good alternative to a new population study was to derive new regression equations from collated good quality measurements, complying with temporal recommended standards; such data were not available. The first use of collated datasets for deriving predicted values for children was based on 6 data sets from 5 European countries [13]. This study showed that the resulting reference values fit 5 of the 6 data sets; it transpired that the sixth set had been affected by a technical problem. Thus this approach was validated; it led to recommending the American Thoracic Society (ATS) and European Respiratory Society (ERS) to support this technique with a view to deriving reference values based on large groups with a wide age range [13].

In 2005 the European tradition of combining standardisation reports with sets of recommended predicted values came to an end: a joint ATS/ERS committee [14] recommended predicted values for the United States and Canada, leaving the rest of the world uncovered. In 2006 one of us (PHQ) started to remedy the deficiency, aiming to cover as large an age range as possible as well as various ethnic groups. In 2008 over 30,000 records had been generously made available from all over the world, and a manuscript was being prepared, but this was suspended because an ERS working group with the same objectives was founded. This group subsequently acquired ERS "Task Force" status in 2010, and the support of 6 large international societies [15].

2008 was also the year of the groundbreaking publication from Stanojevic *et al.* [16], applying a new and very powerful statistical technique on collated spirometric data from whites in the 3-80 year age range.

The collaborative work in the group that was named "Global Lung Function Initiative" [15] was a privilege thanks to the effective and friendly cooperation, based on mutual respect and trust, with some 70 groups from all over the globe. The analytical work was performed by the "Analytical Team" (Fig. 2).

Situation in 2006

Displaying the predicted FEV1 in white males according to 30 different authors (Fig. 3) reveals a quite worrying picture. For the same height and age predicted values may differ by 1 litre or more. Predicted values for children and adolescents are quite disjointed from those for adults. These prediction equations were used in many parts of the world for diagnostic purposes! A worrisome state of affairs.

Modelling lung function

Until very recently regression equations for lung function were based on simple additive linear regression techniques. The by far most popular models had the following form:

$$Y = a + b \cdot height + c \cdot age + error \text{ (adults)}$$
$$\log(Y) = a + b \cdot \log(height) + error \text{ (children)}$$

Fig. 2 - The "Analytical Team" of the "Global Lung Function Initiative". From left to right: Prof. Tim Cole, Prof. Janet Stocks, Prof. Philip Quanjer, Dr Sanja Stanojevic.
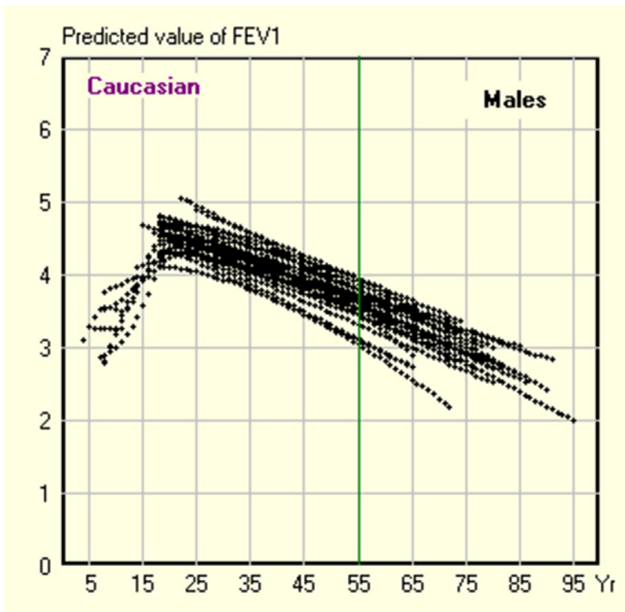
Fig. 3 - Predicted FEV1 in white males. Derived from software download-able from www.spirxpert.com/GOLD.html.

Y is the predicted value, for example FEV1. The "error", also called residual, is the difference between measured and predicted value. For children and adolescents the indices are usually log transformed, and age is rarely taken into account. When using the above linear models it is commonly assumed that the residuals are the same at any combination of age and height.

Fig. 4 displays FEV1 as a function of age in a large number of healthy females aged 3-95 year. It illustrates a few points:
1 The relationship cannot be characterised by straight lines.
2 The scatter ("error") is not constant.
3 The scatter is not proportional to the predicted value.

We can calculate the predicted values for FEV1 for the females in fig. 4 using the widely used ECCS/ERS prediction equations. The mean difference between measured and predicted value of FEV1 should be 0 if the equation fits the data perfectly. Figure 5 shows that there is a systematic difference: the measured FEV1 is on average 180 mL larger than predicted. The values predicted by ECCS/ERS are therefore systematically too low.

Fig. 5 - Difference between measured and predicted FEV1 in healthy white females when using the ECCS/ERS prediction equations.

This brief introduction leads to the following conclusions:
1 The separation of children/adolescents and adults is artificial and leads to disjointed predicted values at the transition from adolescence to adulthood.
2 The models fit the measured values poorly, particularly in children.
3 Differences in predicted values by various authors are very large.

Use of percent of predicted

When interpretating spirometric data, it is an ingrained habit in respiratory medicine to express measured values as percent of predicted. This tradition probably arose from a recommendation by Bates and Christie [17]: "a useful general rule is that a deviation of 20% from the predicted normal value probably is significant". This leads to considering 80% of predicted as the "lower limit of normal" (LLN). This rule of thumb was uncritically adopted. The rule is only valid if the scatter around the predicted value is proportional to that value; hence, large if the predicted value is large, and proportionally smaller if the predicted value is small. As shown in fig. 4 there is no proportionality, so that the use of percent of predicted will inevitably lead to erroneous interpretation of test results (fig. 6), as has been explained
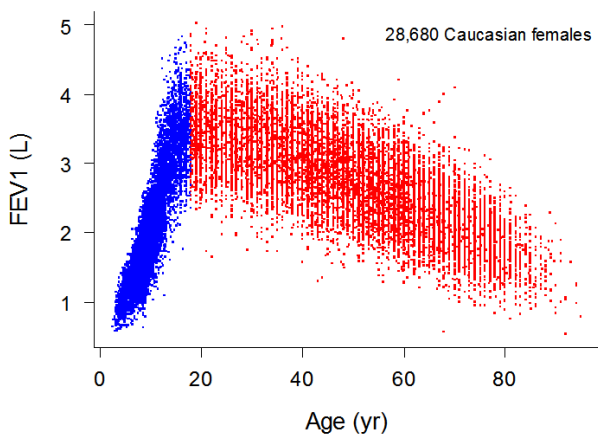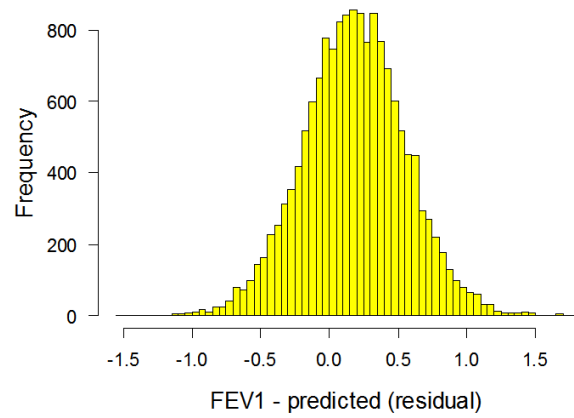

Fig. 4 - Relationship between age and FEV1 in 28,690 white, healthy females. About half of the scatter is due to differences in standing height.
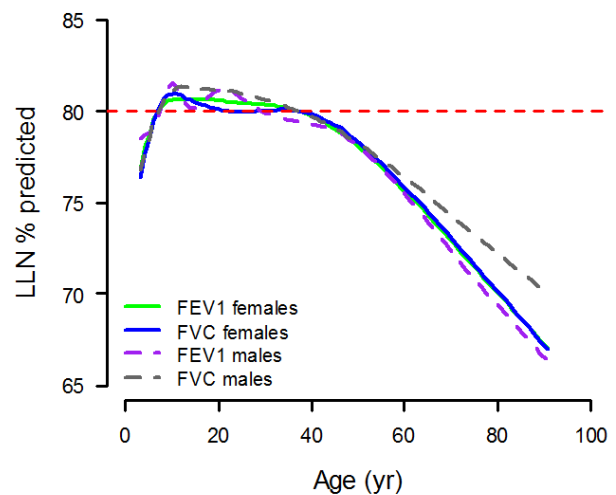

Fig. 6 - The lower limit of normal (LLN) for FEV1 and FVC expressed as a percentage of the GLI-2012 predicted values in the 3-95 year age range.
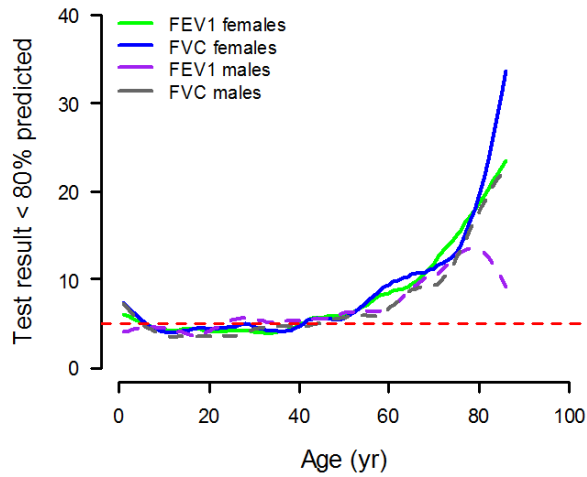
Fig. 7 - Percentage of healthy males and females in whom the measured FEV₁ or FVC is <80% predicted.
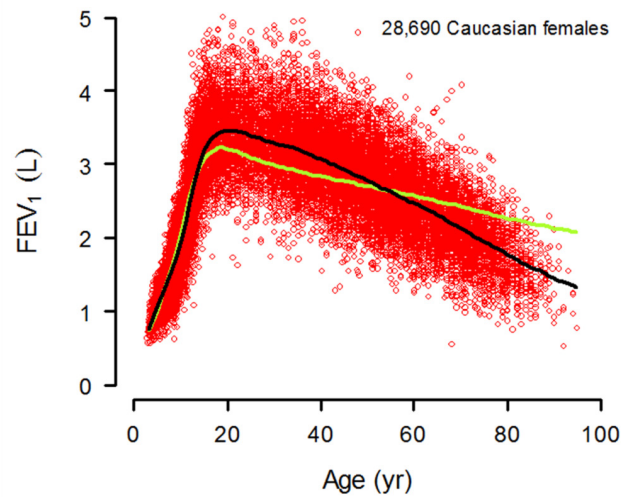

Fig. 9 - The predicted FEV₁ without use of a spline (yellow-green line) provides a bad fit, the one which includes a spline (black line) fits well.

in scores of publications [10,16,18-23]. In fact, Sobol wrote [19]: "Nowhere else in medicine is such a naïve view taken of the limit of normal". As the GLI group had tens of thousands of records available, this provided an opportunity to estimate the LLN more accurately (see later). Expressing the LLN as %predicted leads to the picture in figure 6: over a large age range the LLN is well below the 80% predicted line.

We can subsequently assess in what percentage of a healthy, non-smoking population (25,827 males, 31,568 females) the measured FEV₁ and FVC are below the 80% predicted mark (fig. 7). It will be clear that the large proportion of erroneous assessments of test results, in particular in those aged over 50 years, should lead to abandoning the use of %predicted.

Global Lungs Initiative: what is new?

Capturing the non-linear relationship between spirometric indices and age and height, using standard linear regression techniques, is not possible. Occasionally this predicament was solved by splitting the age range up in two: adults, and children and adolescents, and deriving two sets of equations that joined well, see *e.g.* Hankinson *et al.* [24]. Prior to that childhood was covered by a more complex model [13],
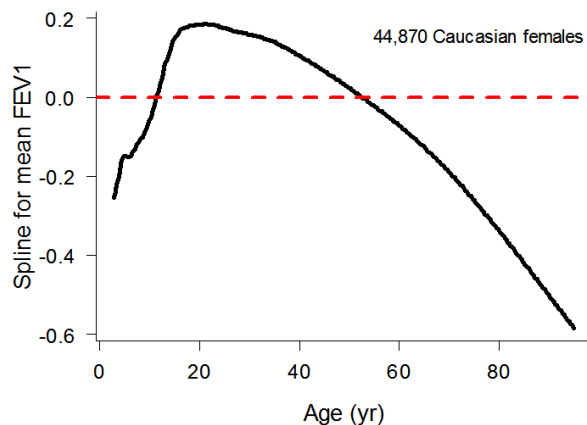

Fig. 8 - The "spline", which adds an age-specific term to the predicted value. Please note that the prediction equations use a logarithmic scale.

or by a large number of regression equations, each spanning one year [25]. More sophisticated models were used similarly for the adult age range, paying special attention to accurately defining the LLN [26-27]. An elegant method for capturing non-linear curves is by adding a "spline" to a linear relationship:

$$\log(Y) = a + b \bullet \log(height) + c \bullet \log(age) + spline + error$$

This approach was adopted by Pistelli *et al.* [28-29]. However, the statistical package GAMLSS [30], first used to this end by Stanojevic *et al.* [16], offers more advanced methods for modelling pulmonary function. In practice the spline is modelled as a function of age. You can best envisage this as an age-specific adjustment of the predicted value: a correction that varies with age in the 3-95 year age range (figure 8). We operate on a logarithmic scale. This implies, *e.g.* in a 20 year old woman, that the predicted FEV₁ calculated using the linear coefficients (a, b and c in the above equation) should be multiplied by exp(0.19) = 1.21, hence a 21% increase. In a 85 year old women we multiply by exp(-0.40) = 0.67, correcting the FEV₁ by 33%.

The difference between the predicted value with and without spline is illustrated in figure 9. The yellow-blue line represents the predicted value without spline. In children and adolescents the fit looks passible, but in adults the fit is very poor. Conversely, the black line, which represents the predicted value when adding a spline function, fits the actual values over the entire age range.

FEV₁/FVC: a surprise

Analysis of the FEV₁/FVC ratio led to an unexpected result. The predicted value fell quickly between 3 and approximately 10 years of age, followed by a small increase up to about 16 year, and then a gradual non-linear decline in adults (figure 10). As this pattern had never been described before the first thought was that we were dealing with an artefact arising from the collation of so many datasets. After all, if one centre would contribute data with an unusually low FEV₁/FVC ratio in and around the 10 year age range, this could explain the findings. However, no centre had contributed
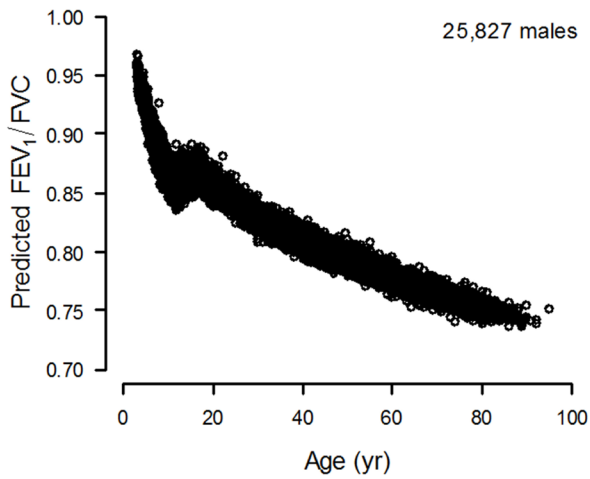
Fig. 10 - Predicted FEV₁/FVC in white males.

a group of children limited to this fairly narrow age range.

Evidence that the findings were not an artefact came from analysis of data from boys and girls from 15 different centres, comprising different ethnic groups (figure 11, [31]). As the determinants of the FEV₁ and the VC are not the same, it follows that after birth the vital capacity grows proportionally faster than the FEV₁, and that this pattern is temporarily reversed during the adolescent growth spurt [31].
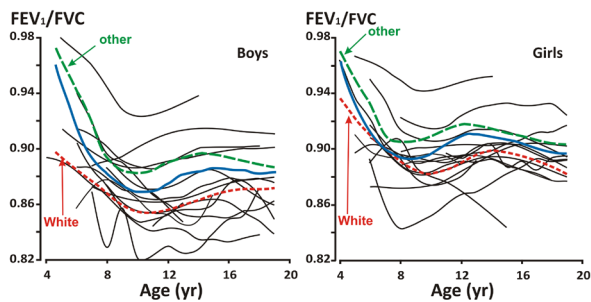

Fig. 11 - Data from 15 centres, comprised of different ethnic groups, all display the same pattern: rapid decline of FEV₁/FVC ratio until the start of the adolescent growth spurt, then a small increase followed by a decline.

"Lower limit of normal"

In clinical medicine, the 'normal range' is generally defined as the range of values which encompasses 95% of a healthy population. The lower limit of normal (LLN) is the cut-off below which results from only 2.5% of healthy individuals will fall, while the upper limit of normal (ULN) represents the threshold above which results from only 2.5% of healthy individuals will be found. Accordingly 95% of the healthy population is considered to have
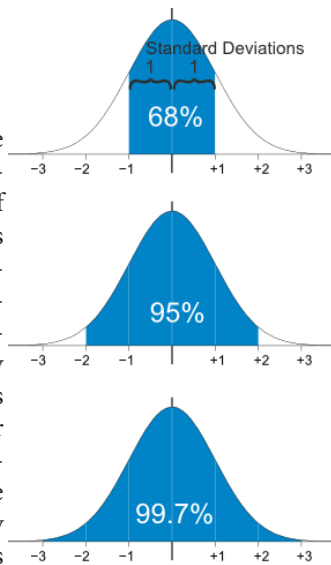

Fig. 12 - Relationship between standard deviation and percentage of data under the curve in the case of a normal distribution.

"normal" test results, whereas in 2½% they are "too low" and in 2½% "too high", resulting in 5% false-positive test results. Results of spirometric tests characteristically lead to values for FEV₁ and VC which are too low rather than too high in disease. This probably explains why in respiratory medicine the LLN is defined as that value which identifies the lower 5th centile of a healthy population of non-smokers.

There are various methods for technically deriving the LLN. The most elegant one is based on a "normal distribution" of test results. In that case (fig. 12) 68% of observations are between +1 and -1 standard deviation (SD) of the distribution, 90% between +1.64 and -1.64 SD, 95% between +1.96 and -1.96 SD, and 99.7% between +3 and -3 SD.

In a healthy subject spirometric data vary with age, height, sex and ethnic group. After taking these into account we are left with the residual (measured – predicted value). If the residual is normally distributed the average of residuals is 0. Dividing the residuals by the SD of the distribution [(measured - predicted)/SD] yields a dimensionless number, the z-score. In the case of a normal distribution the average of all z-scores is 0, and the SD is 1 (fig. 12).

The SD (or coefficient of variation: CoV = 100•SD/predicted) varies with age [16,23]. Hence the CoV must be modelled so that we obtain a normal distribution, *i.e.* independent of age. Again a spline can be used for optimal modelling:

$$\log(CoV) = a + b \cdot \log(age) + spline + error$$

The coefficient of variation for FEV₁ in white females varies between 12½% and 25% (fig. 13). How does this affect the LLN? At ages 3, 20 and 80 year the CoV is approximately 16%, 12½% and 21%, respectively. The LLN in respiratory medicine is the 5th percentile, when the z-score is -1.64, *i.e.* at the predicted value minus 1.64 times the CoV. It follows that the LLN for FEV₁ in a 3, 20 and 80 year old white healthy female is at 74%, 80% and 66% of the predicted value.

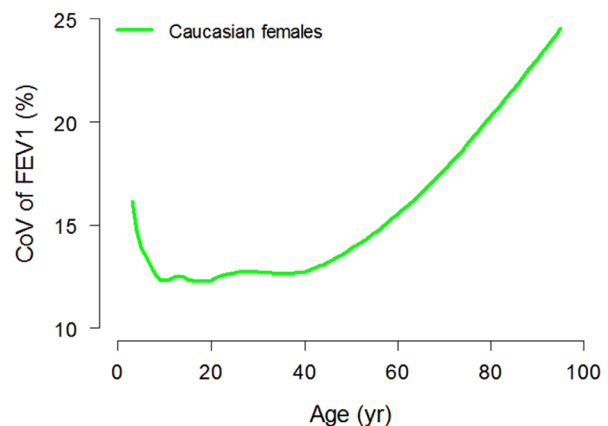Once again confirmation that we should not regard 80% of predicted as the LLN.

Fig. 13 - The coefficient of variation (CoV) for FEV₁ in healthy white females varies with age.
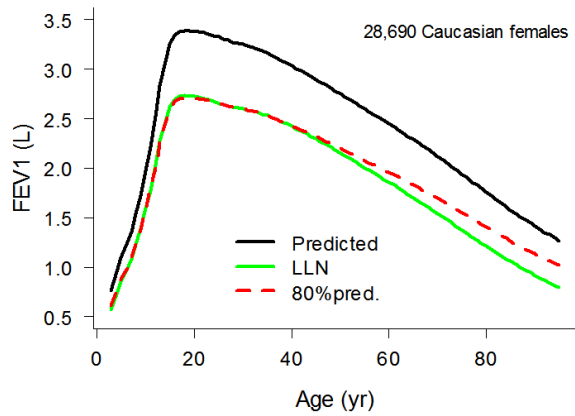
Fig. 14 - Predicted FEV1 and LLN in healthy white females, and 80% predicted, as a function of age.

Fig. 16 - FEV1/FVC ratio in healthy females of different ethnic origin.

To put this in further perspective we can depict the predicted value and the LLN for FEV1 (according to GLI-2012) in white females as a function of age (fig. 14). Adding the line representing 80% of predicted illustrates that, particularly in adults, this line creeps progressively higher up in the normal range, leading to a progressively larger proportion of false-positive test results.

As explained above the procedure adopted should lead to a normal distribution of residuals, so that the z-scores have an average of 0 and SD 1. Figure 15 demonstrates that this is achieved with the statistical package GAMLSS. This is associated with tremendous benefits: the z-score is completely independent of age, height and sex. For example, if the z-score for any index is -1.64, this signifies in males, females, children and adults that the measured value is at the 5th percentile; in lung function testing this is regarded as the LLN.
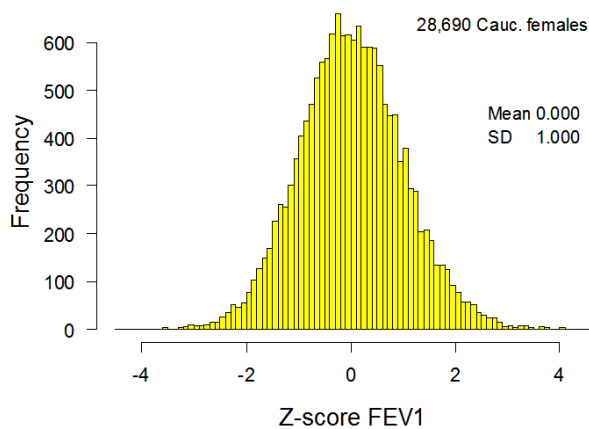

Fig. 15 - Distribution of z-scores for FEV1 in healthy white females.

Ethnicity

It is well known that pulmonary function differs between ethnic groups. In the past one used "ethnic correction factors", implying that predicted values for a pulmonary index of, for example, black subjects were calculated as being about 15% below those of whites. These "correction factors" were determined empirically in adults. The availability of a large
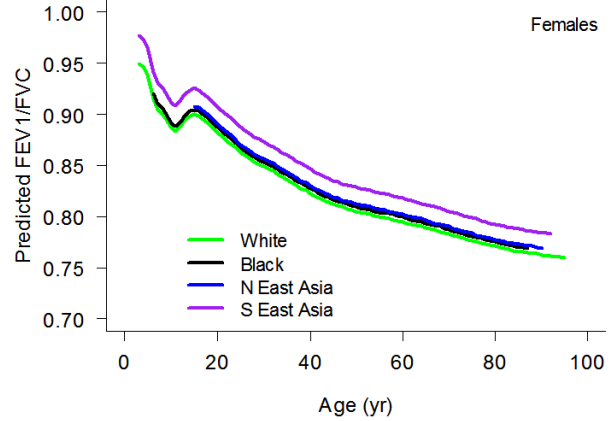
number of spirometric records from 3-95 year old subjects of different ethnic background allowed the Global Lung Function Initiative to look into ethnic differences in greater depth. Fig. 16 illustrates an important observation: with the exception of South East Asians (southern China, Thailand, Korea), the FEV1/FVC ratio is the same in all ethnic groups at any given age and height. This implies that differences in FEV1 and FVC between ethnic groups are proportional, and independent of age. Biologically this makes sense. After all, all ethnic groups belong to the genus *Homo sapiens*, *i.e.* mammals comprising subgroups that adapted to different local conditions and differ in socio-economic backgrounds. In an evolutionary process covering millions of years mammals have been provided with a scalable lung design; as it is scalable it fits small and large animals, catering for their metabolic and other needs under widely different circumstances [32]. Differences in pulmonary indices between ethnic groups are therefore no more than a matter of different scale. Based on this finding of proportional differences we can now add ethnic group to our model, as follows:

$$\log(Y) = a + b \bullet \log(\text{height}) + c \bullet \log(\text{age}) + d \bullet \text{Ethn} + \text{spline} + \text{error}$$

Ethnicity (Ethn) is now a co-factor. Mean differences in pulmonary function of a number of ethnic groups, relative to whites, are shown in table 1. A group "Mixed/other" denotes people of mixed ethnic descent; the figures in the table are an estimate, pending further studies.

The above represents an important step forward, as all ethnic groups can now be included in the regression equation. This does not solve all problems, as there appear to be differences in the scatter around predicted values. This implies

Table 1 - Percentage difference in pulmonary function, by sex and ethnic group, compared to whites [23]

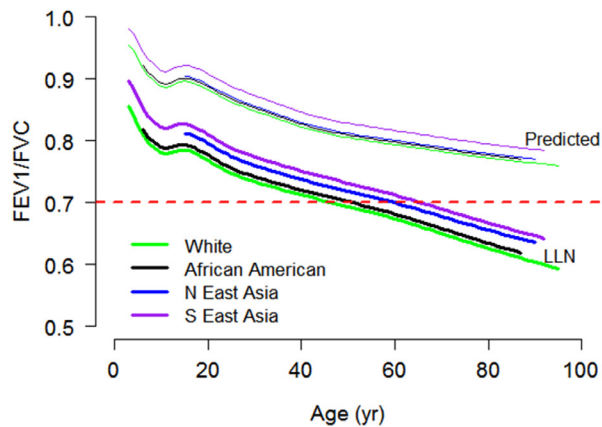|	Females			Males		
	FEV1	FVC	FEV1/FVC	FEV1	FVC	FEV1/FVC
African-American	-13.8	-14.4	0.6	-14.7	-15.5	0.8
North East Asian	-0.7	-2.1	1.1	-2.7	-3.6	0.9
South East Asian	-13.0	-15.7	2.9	-9.7	-12.3	2.8
Mixed/other	-6.8	-7.9	1.1	-6.8	-7.9	1.1

Fig. 17 - Predicted FEV1/FVC ratio and lower limit of normal (LLN) in healthy females of different ethnicity.

that it is necessary to adjust the model for the coefficient of variation, shown earlier, as follows:

$$\log(CoV) = a + b \bullet \log(age) + d \bullet Ethn + spline + error$$

The FEV1/FVC ratio is a pivotal objective index for diagnosing pathological airway obstruction. Whereas the predicted values for this ratio differ scarcely between ethnic groups, the LLN is clearly different (fig. 17). The GOLD group considered it too difficult to calculate the LLN for the FEV1/FVC ratio and decided that it was much easier to adopt a fixed LLN of 0.70. A lot of criticism has been published about the unscientific approach and the lack of any evidence that obstructive lung disease can thus be properly diagnosed. See for example an Open Letter, signed by a large number of reputable researchers and clinicians [33]. Figure 17 also discloses that the GOLD criterion might lead to the spurious finding that COPD is less prevalent in East Asians, as the LLN for FEV1/FVC remains above the 0.70 limit until a higher age than in whites and blacks.

The "lower limit of normal" once more

There can be little doubt that the distribution of lung function indices of healthy subjects and those with lung pathology overlaps. It is therefore risky to conclude that a test result > LLN excludes pathology; it goes without saying that clinical judgement matters. On that account it has been suggested that a FEV1/FVC ratio < 0.70 but > LLN, hence within the normal range and dubbed the "twilight zone", represents lung pathology. Evidence to support this is lacking. However, if subjects in the "twilight zone" develop respiratory symptoms and signs after a number of years, this might lend support to this claim. Supportive evidence has not been found in longitudinal studies:

GOLD stage 1 (FEV1/FVC < 0.70 & FEV1 > 80%) in asymptomatic subjects is **not** associated with
- Premature death [34-38]
- Accelerated decline in FEV1, development of respiratory symptoms, increased use of health care, decrease in

"quality of life" [39].

FEV1/FVC < LLN **is** associated with
- Premature death [35,40]
- Development of respiratory symptoms [41].

Conclusion: The GOLD criterion is unscientific, clinically unfounded, and the use of FEV1/FVC < 0.70 as a criterion for diagnosing airway obstruction should be discouraged in view of under diagnosis in young subjects and extensive over diagnosis in elder adults [33].

Ethnicity and z-score

It does not do any harm to illustrate the usefulness of the z-score from yet another perspective. Going from left to right in fig. 15, the z-scores relate to an ever increasing proportion of the population. Replace the absolute count with the cumulative percentage of the population on the Y-axis and you get fig. 18. The scale is from 0 (0 subjects) to 1 (all subjects covered, 100% of the population). The cumulative frequency distribution of white females is indistinguishable from that of black females (fig. 18). This illustrates once more the great utility of z-scores, as they can be interpreted independent of ethnic group.

Interpretation of test results

Lung function tests produce a once-only result. The result does not only reflect the presence or absence of respiratory disease, but is also influenced by the time of the day, daily and seasonal variation, *etc*. (fig. 19). Such spontaneous variability should always be taken into account when interpreting test results [42].

The way in which spirometric test results are usually presented does little to facilitate interpretation and mystifies the inexperienced assessor: observed values of FEV1, FVC, FEV1/FVC together with additional indices, such as pre and post bronchodilator, predicted values, lower limits of normal, percent of predicted, represents an impenetrable array of data that confuses most recipients, whether clinicians, technicians or patients. Conversely, pictograms in which z-scores are depicted relative to a normal range allow
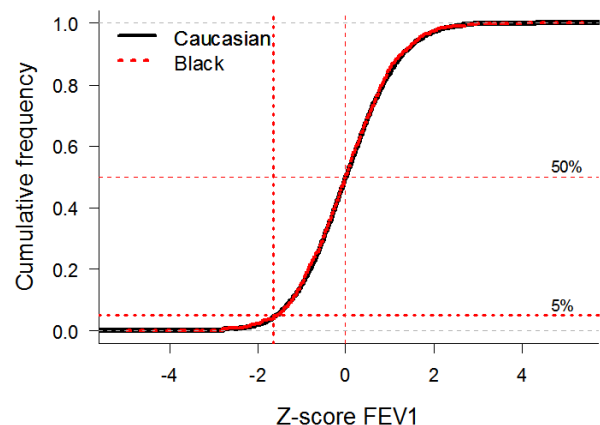

Fig. 18 - Cumulative frequency distribution of z-scores for FEV1 in healthy non-smoking white and black females.
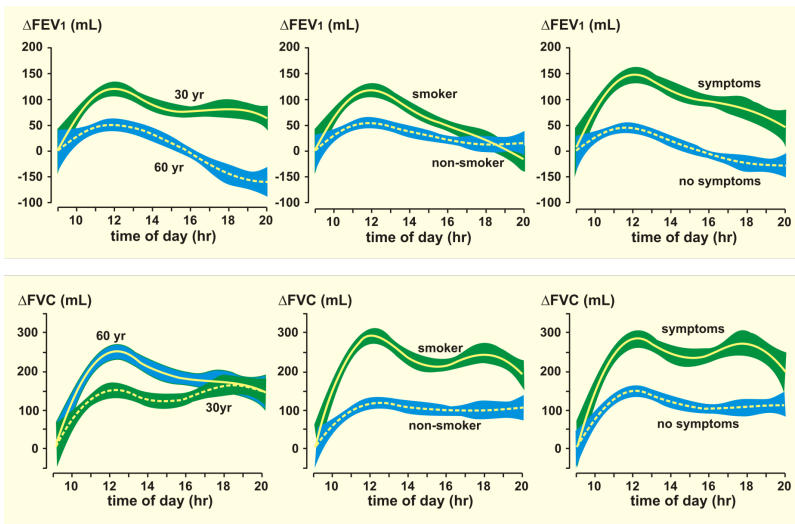
Fig. 19 - Circadian and seasonal variation in the level of pulmonary function. Data derived from a normal population, from measurements made at 3 year intervals for up to 12 years [42].
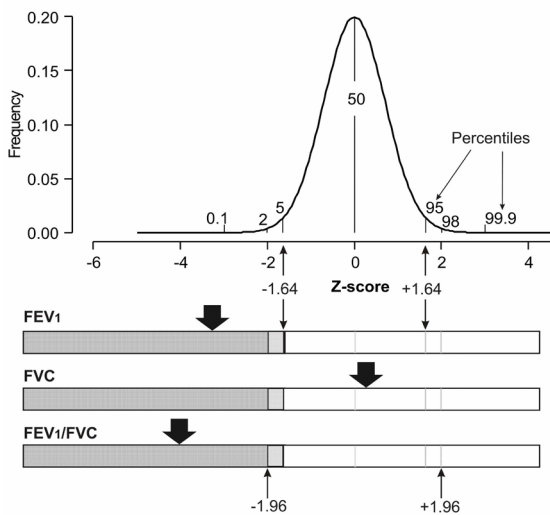
interpreting the findings in the wink of an eye (fig. 20 and 21).

Comparison of predicted values

Paediatricians in the Netherlands rely al-most exclusively on predicted values from Zapletal [43]. These are based on a quite limited number of children (111 boys and girls), and the regression equations only take height into account, not age (6-17 year). In other countries predicted values from Polgar [43], Knudson [44], Quanjer [13], Rosenthal [45], Wang [46] and Hankinson [24] are frequently used. Predicted values according to Stanojevic [16] fit a population of healthy children well, unlike those from Zapletal, Polgar, Wang, Rosenthal, Knudson (fig. 22).

In adults (fig. 23) the FEV1/FVC ratios accord-ing to ECCS/ERS [10] and NHANES [24] dif-fer from those of GLI-2012 [23]. This is mainly due to the fact that the GLI-2012 equations take into account that the ratio is inversely related to standing height, whereas the two other equa-tions only take age into account. Predicted val-ues for FEV1 and FVC according to NHANES agree well with those from GLI-2012, the ECCS/ERS predicted values are definitely too low (fig. 24). Consequently, the ECCS/ERS predicted values, which are widely used in Europe, need to be abandoned.

Fig. 20 - Relationship between percentile and z-score, and its use in a pictogram to facilitate the in-terpretation of test results.

Fig. 21 - The large number of data is not conducive to an easy interpretation of lung function measurements. The use of pictograms, which summarise the findings (bottom left), enables interpretation at a glance.
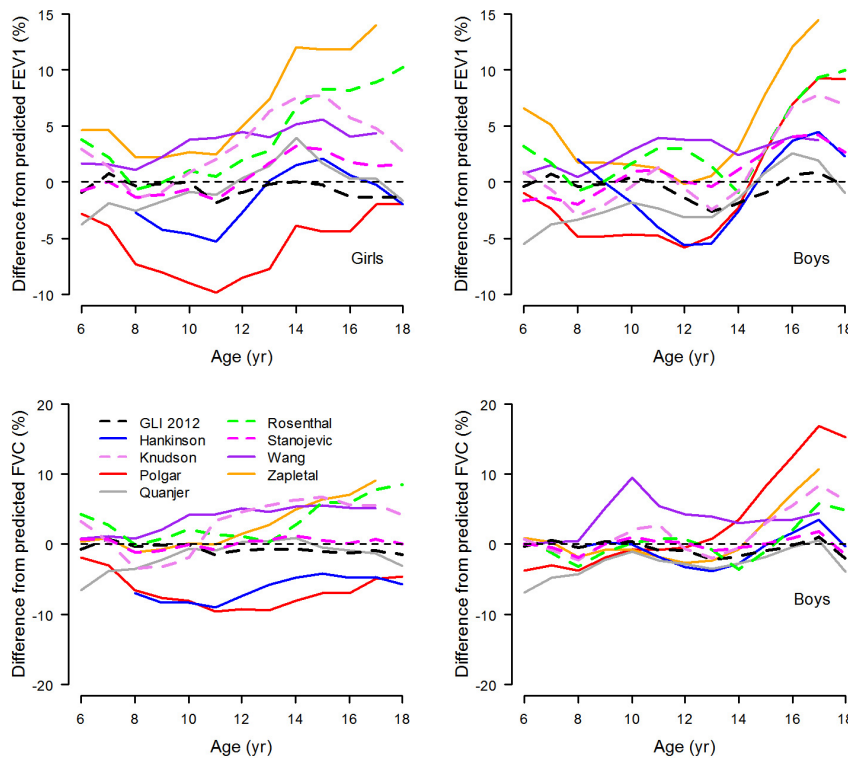
Fig. 22 - Comparison of predicted FEV1 and FVC in healthy boys and girls according to GLI-2012 [23], Zapletal [43], Stanojevic [16], Polgar [12], Quanjer [13], Hankinson [24], Knudson [44], Rosenthal [45] and Wang [46].

of the Quanjer GLI-2012 equations will not lead to a clinically significant change in the prevalence rate of airway obstruction.

As explained earlier GOLD stage 1 is not regarded as representing lung disease. Therefore the analysis is limited to GOLD stages 2-4 (fig. 27). The prevalence rate of GOLD stages 2-4 has the same pattern as previously published for GOLD stage 1 (fig. 28): under diagnosis (~20%) of airway obstruction up to age 55-60 year, and over diagnosis (~20%) above that age. These percentages agree with those reported in an earlier clinical study [47]. This indicates that an age-related bias even affects GOLD stage 2. This is in part due to the fact that the FEV1 should be < 80% of the predicted value. We concluded earlier that not only FEV1/FVC < 0.70 (fig. 17), but also FEV1 < 80%, was associated with a strong age-related bias (fig. 6, 7 and 14).

Airway obstruction

Applying predicted values for FEV1/FVC according to various authors on data from paediatric patients from the Children's Hospital of Pittsburgh (courtesy Dr. Weiner) discloses differences in the prevalence rate of airway obstruction in boys, less so in girls (table 2).

Data a wide ranges of diagnoses from two hospitals in Australia and one in Poland (fig. 25) disclosed the following trend (fig. 26). There is fair agreement in the prevalence rate of airway obstruction according to GLI-2012 and NHANES predicted values, with NHANES in women producing a systematically higher prevalence rate. The ECCS/ERS prediction equations (fig. 27) lead to a somewhat lower prevalence rate in males up to 60 year, and in young females. In general differences are relatively small; hence adoption

"Restrictive pattern"

In 1991 an ATS-committee suggested that it was possible to uncover a restrictive ventilatory defect, *i.e.* a condition in which the total lung capacity is reduced, on the basis of

Table 2 – Prevalence rate of airway obstruction according to GLI-2012 and other prediction equations.

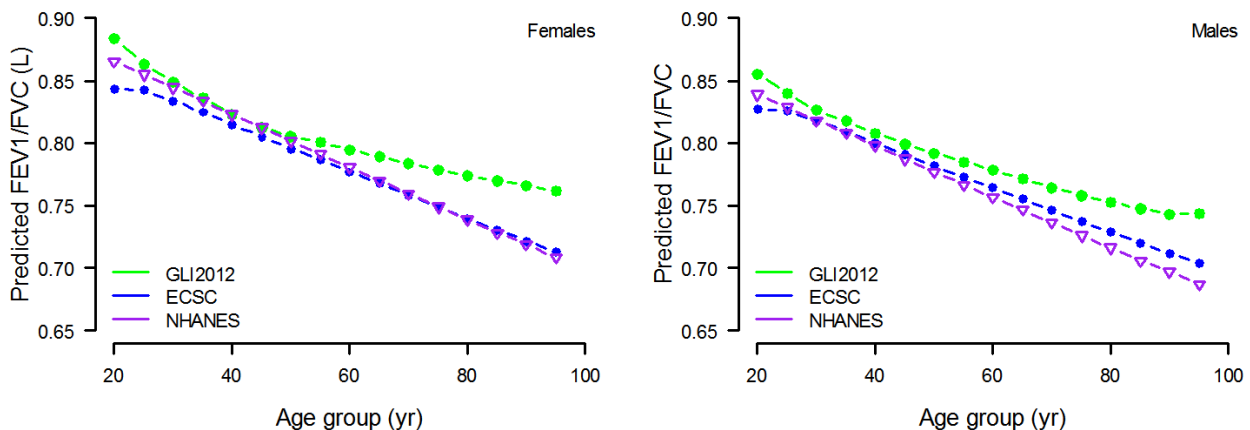|	FEV1/FVC < LLN	
Author	**Boys** n = 2492	**Girls** n = 2072
Hankinson	17.8%	14.3%
Knudson	21.0%	10.5%
Quanjer GLI-2012	15.0%	14.0%
Wang	21.6%	16.8%
Zapletal	23.1%	10.9%

Fig. 23 - Comparison of predicted FEV1/FVC ratio in boys and girls according to GLI-2012 [23], Hankinson[24] and ECCS/ERS [10].
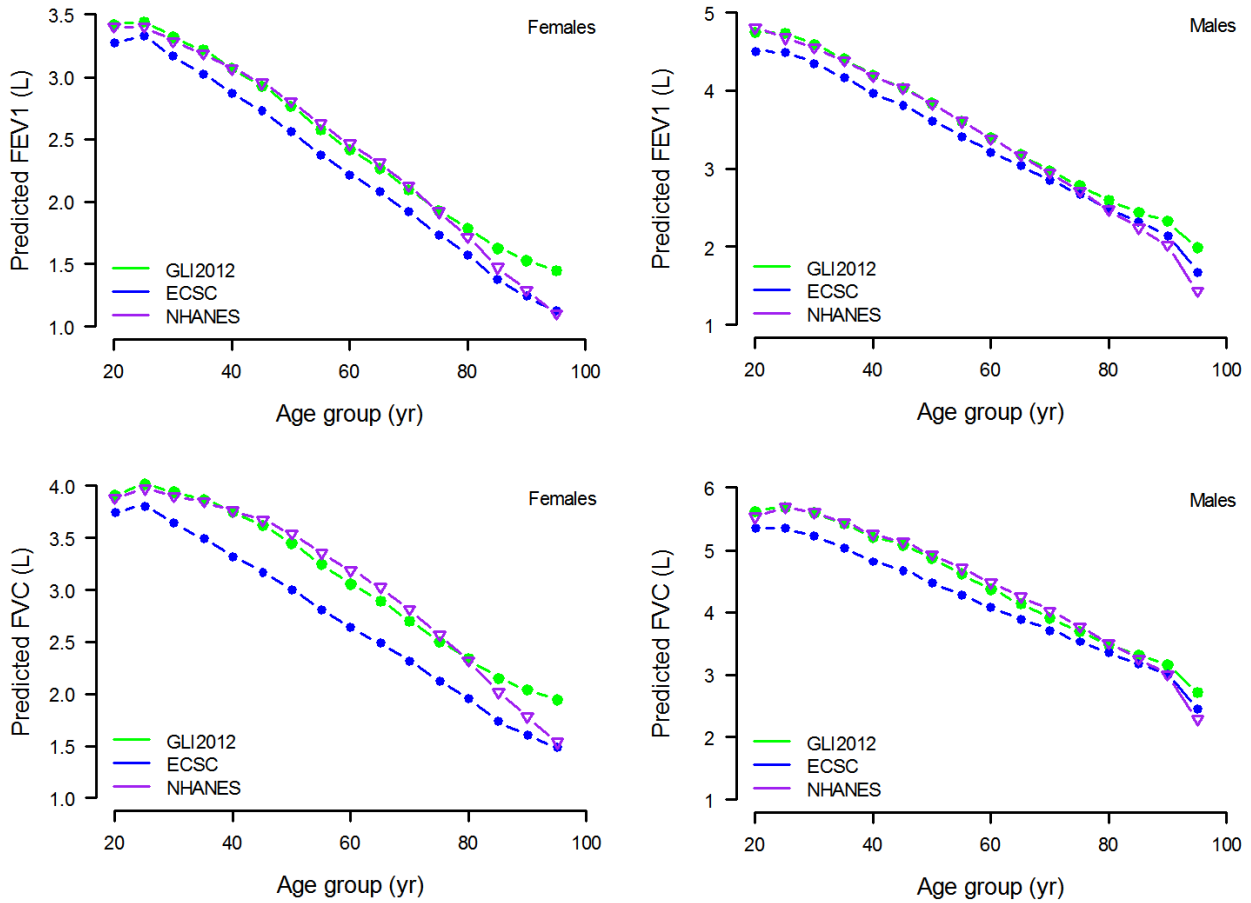
Fig. 24 - Comparison of predicted FEV1 and FVC in healthy adults according to GLI-2012 [23], ECCS/ERS [10] and NHANES [24].
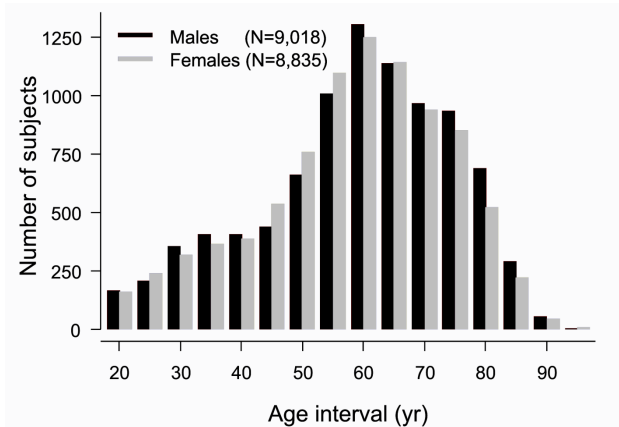

Fig. 25 - Age distribution of patients (Australia, Poland).

an abnormally low VC in combination with a normal or high FEV1/FVC ratio: "restrictive pattern" [21]. Since then a restrictive pattern has been regularly described in the literature, suggesting that it is considered a clinically meaningful pattern. The prevalence rate in an Australian and Polish population of hospital patients (fig. 25) varied with age between 5 and 20% (fig. 29); the number of observations above age 80 year was very limited, so that the pattern above that age should be neglected. Differences in the prevalence rate according to the three sets of prediction equations are considerable. The general pattern is that adopting the GLI-2012 equations leads to an increase in the prevalence rate of a restrictive pattern compared to ECCS/ERS. This is worrisome, as it may lead to an increase in requests
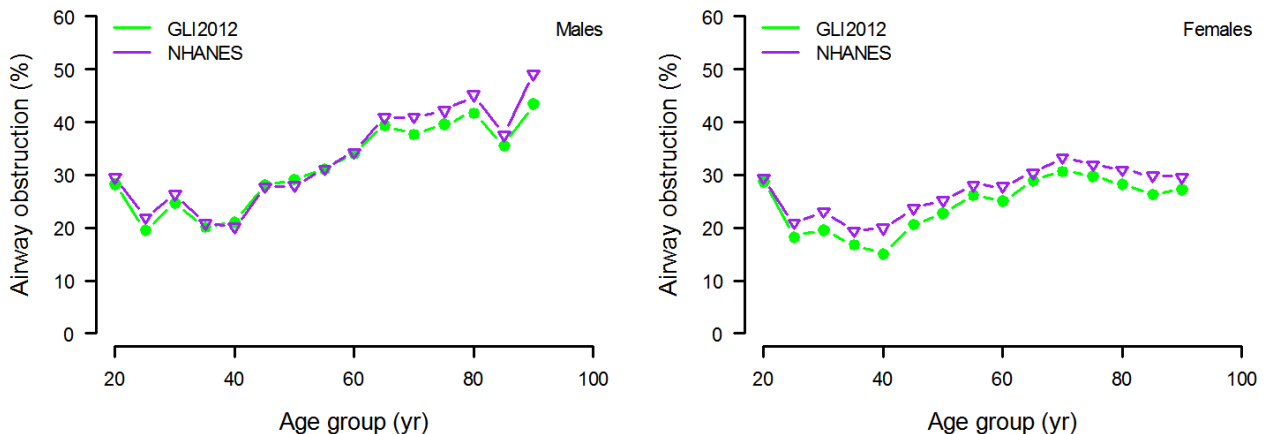

Fig. 26 - Percentage of patients with airway obstruction (FEV1/FVC < LLN) based on GLI-2012 [23] and NHANES [24] prediction equations.
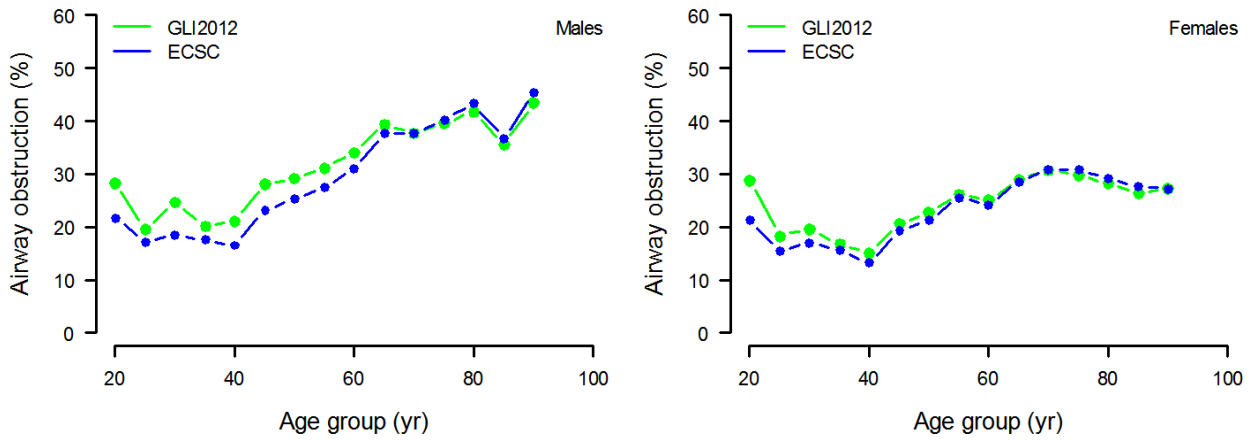
Fig. 27 - Percentage of patients with airway obstruction (FEV1/FVC < LLN) based on GLI-2012 [23] and ECCS/ERS [10] predicted values.
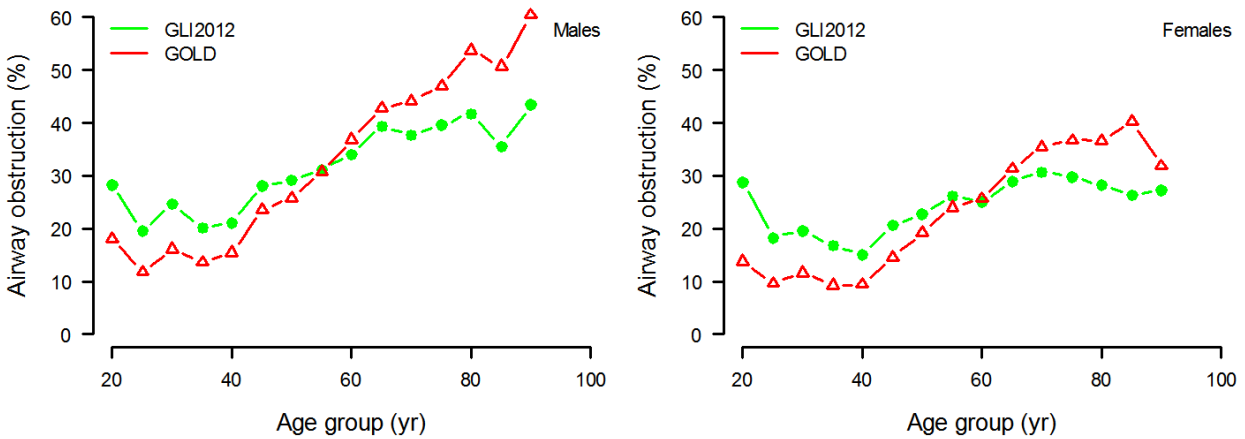

Fig. 28 - Percentage of patients with airway obstruction (FEV1/FVC < LLN) based on GLI-2012 [23] predicted values, or with GOLD stage 2-4.

to measure the total lung capacity, leading to an increase in medical expenditure. It is known that this spirometric pattern has a low sensitivity for correctly diagnosing restrictive lung disease: 50% or less in a clinical population [48-50]. Lung restriction is rare in the general population, so that it is best if general practitioners ignore a restrictive pattern. In fact, in general it is better to ignore this pattern, unless there is clinical evidence compatible with lung restriction (lung resection, severe kyphoscoliosis, *etc.*) and documenting such a defect is clinically relevant. The general idea should be: "treat the patient, not the numbers".

Accurate measurement of height and age
Height
Height should be measured, as self-reported height is unreliable. Differences between actual and self-reported height may be up to 6.9 cm, and are generally largest in elderly subjects [51-56]. The FEV1 and FVC are a function of height^k, where k ~ 2.2. In a 110 cm tall child, or a 180 cm tall adult, a 1 cm error leads to an error in the predicted lung function index of 2% and 1.2%, respectively. Not only should standing height be measured, but the stadiometer should be calibrated every year, and in
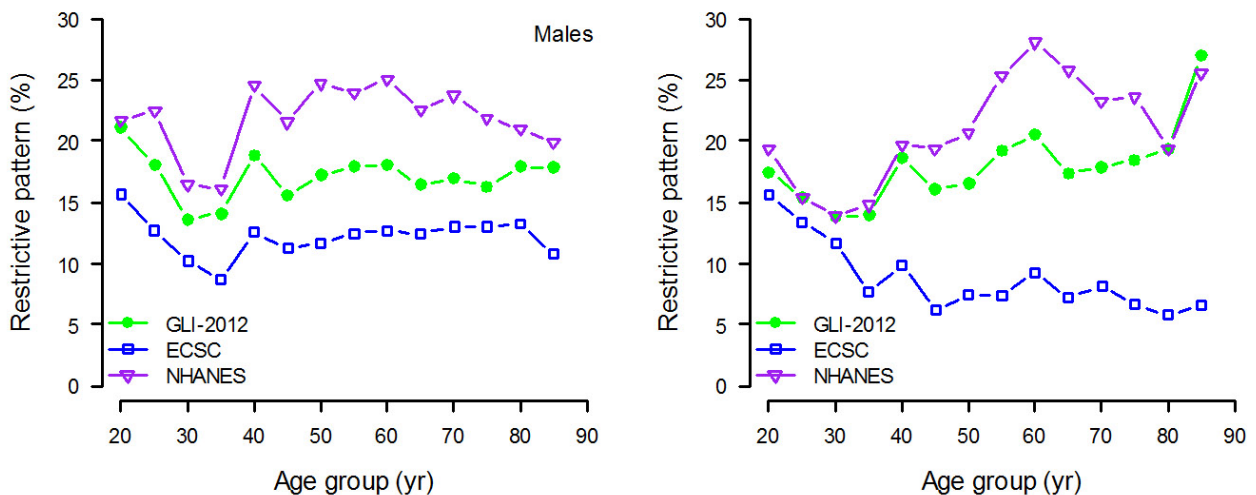

Fig. 29 - Percentage of patients with a spirometric "restrictive pattern": VC too small but normal or high FEV1/FVC ratio.

calculating predicted values height should be entered with 1 decimal accuracy [23, 57].

Age

The effect of errors in age on predicted values cannot be so easily estimated because of the variable contribution of the spline in age. If age is systematically underestimated by 0.75 years by rounding off, then the percentage error is as listed in table 3.

Table 3 - Rounding off age, here by 0.75 year, leads to errors in the predicted values for FEV_1 and FVC.

	Males		Females	
Age (yr) (rounded off)	FEV_1 % error	FVC % error	FEV_1 % error	FVC %rror
3 *vs* 3.75	-2.8	-3.4	-2.9	-3.6
10 *vs* 10.75	-1.3	-1.4	-2.6	-2.7
15 *vs* 15.75	-3.4	-2.9	-3.4	-2.9
50 *vs* 50.75	+0.4	+0.4	+0.6	+0.7
85 *vs* 85.75	+0.7	+0.5	+0.9	+1.0

The errors vary with age, the largest errors occurring in childhood. Therefore, in calculating predicted values, age should be entered with 1 decimal accuracy [23, 57].

Validation

The GLI-2012 predicted values have been validated in 2 studies [58-59].

Software

Two kinds of (free) software are available to generate predicted values according to the Quanjer GLI-2012 reference equations:

1 **Software for calculating predicted values for an individual** This software is available as a desktop program for Windows systems, and in the form of an Excel spreadsheet.
2 **Software for transforming large datasets** so that predicted values, LLN and z-scores are added to the data. This free software is similarly available as a desktop application for Windows systems, and as an Excel spreadsheet.
The software can be downloaded from here.

In addition spirometer manufacturers have implemented the GLI-2012 equations in their software, or are in the process of doing so. Information is to be found at this location.

Flows

There are recurrent questions why predicted values for instantaneous flows, such as FEF_{50}, have not been included in the GLI-2012 set. These flows have never been shown to have added value over and above FEV_1 and VC. These flows are often considered to be sensitive indices of "small air-ways disease", a syndrome that would occur without affecting large intrapulmonary airways in a manner that would be detectable by spirometry; this view has been contested as early as 1991 [21]. The coefficient of variation of instantaneous flows is quite large, which partly explains their unsatisfactory performance in clinical decision making. Also, flows pre and post bronchodilator cannot be compared if a change occurs in the FVC, or in the case of spontaneous changes in the FVC, and predicted values for flows are invalid if the FVC is affected by the disease process. It is for this reason that the use of instantaneous flows for diagnostic purposes is not recommended in standardisation reports, and that they do not feature in diagnostic algorithms [10,14,21,60].

In paediatrics instantaneous flows are still frequently used. For this reason, at special request, the GLI group added predicted values for $FEF_{75\%}$ and $FEF_{25-75\%}$.

Transfer factor

The GLI group has started deriving predicted values for transfer factor. The group, under the leadership of Brian Graham and Graham Hall, received "task force status" from the ATS.

Transfer factor of the lung is often called diffusion capacity of the lung. However, the lung does not diffuse. In addition the measurement does not represent a capacity, because for example during exercise gas transfer of O_2 or CO across the lung is much greater than during rest. Therefore transfer factor is a better name.

Lung volumes

At this stage there are no plans to derive regression equations for lung volumes (RV, TLC, FRC). This is in part because there are so many different techniques to measure lung volumes, and because few data on healthy subjects are available. In addition many hold the view that the measurement of lung volumes is of limited value in clinical practice.

Conclusions

1 The study performed by the Global Lung Function Initiative is based on a very large and representative population sample.
2 The recommendations have been endorsed by 6 large international respiratory societies: ERS, ATS, Australian and New Zealand Society of Respiratory Science, Asian Pacific Society for Respirology, Thoracic Society of Australia and New Zealand, and the American College of Chest Physicians.
3 GLI-2012 provides regression equations for the 3-95 year age range, and for a number of ethnic groups.
4 The age dependence of the LLN has been accounted for.
5 Z-scores offer the opportunity to interpret test results independent of age, height, sex and ethnic group.
6 Adoption of the Quanjer GLI-2012 equations will lead to minor changes in the prevalence rate of airway obstruction in clinical populations.

7 The use of percent of predicted values leads to an unacceptable age bias and needs to be replaced by the use of z-scores.

8 The GOLD doctrine does not respect the clinically valid LLN and leads to considerable under and over diagnosis of airway obstruction.

9 Adopting the Quanjer GLI-2012 equations will lead to an increase in the prevalence rate of a 'restrictive pattern' compared tot ECSC: "treat the patient, not the data".

Acknowledgements

References

1 Hutchinson J. On the capacity of the lungs, and on the respiratory functions, with a view of establishing a precise and easy method of detecting disease by the spirometer. *Med Chir Trans (London)* 1846; 29: 137–252.

2 Tiffeneau R, Pinelli A. Air circulant et air captif dans l'exploration de la fonction ventilatrice pulmonaire. *Paris Méd* 1947; 37: 624–628.

3 Yernault JC. The birth and development of the forced expiratory manoeuvre: a tribute to Robert Tiffeneau (1910–1961). *Eur Respir J* 1997; 10: 2704–2710.

4 Jouasset D. Normalisation des épreuves fonctionnelles respiratoires dans les pays de la Communauté Européenne du Charbon et de l'Acier. *Poumon Coeur* 1960; 16: 1145–1159.

5 Cara M, Hentz P (1971). Aide-mémoire of spirographic practice for examining ventilatory function, 2nd edn. (Industrial Health and Medicine series, vol 11) pp. 1-130.

6 Ferris BC: Epidemiology Standardization Project. *Am Rev Respir Dis* 1978; 118 (Suppl, part 2): 1-120.

7 American Thoracic Society. 1979. Standardization of spirometry. *Am Rev Respir Dis* 1979; 119: 831–838.

8 Quanjer PH, ed. Standardized lung function testing. Report Working Party Standardization of Lung Function Tests. European Community for Coal and Steel. *Bull Eur Physiopathol Respir* 1983; 19: Suppl. 5, 1–95.

9 American Thoracic Society. Standardization of spirometry: 1987 update. *Am Rev Respir Dis* 1987; 136: 1285–1298.

10 Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault J-C. Lung volume 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. Erratum *Eur Respir J* 1995; 8: 1629.

11 American Thoracic Society. Standardization of spirometry, 1994 update. *Am J Respir Crit Care Med* 1995; 152: 1107–1136.

12 Polgar, G, Promadhat V. Pulmonary function testing in children: techniques and standards. Philadelphia, WB Saunders C, 1971.

13 Quanjer PH, Borsboom GJ, Brunekreef B, Zach M, Forche G, Cotes JE, Sanchis J, Paoletti P. Spirometric reference values for white European children and adolescents: Polgar revisited. *Pediatr Pulmonol* 1995;19: 135-142.

14 Miller MR, Hankinson J, Brusasco V, *et al*. ATS/ERS Task Force. Standardisation of spirometry. *Eur Respir J* 2005; 26: 319-338.

15 http://www.lungfunction.org.

16 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.

17 Bates DV, Christie RV. (1964). Respiratory Function in Disease, p. 91. Saunders, Philadelphia and London.

18 Sobol BJ. Assessment of ventilatory abnormality in the asymptomatic subject: an exercise in futility. *Thorax* 1966; 2: 445-449.

19 Sobol BJ, Sobol PG. Editorial. Percent of predicted as the limit of normal in pulmonary function testing: a statistically valid approach. *Thorax* 1979; 34: 1-3.

20 Miller MR, Pincock AC. Predicted values: how should we use them? *Thorax* 1988; 43: 265-267.

21 ATS Statement. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Resp Dis* 1991; 144: 1202-1218.

22 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.

23 Quanjer PH, Stanojevic S, Cole TJ *et al*. and the ERS Global Lung Function Initiative. Multi-ethnic reference values for spirometry for the 3-95 years age range: the Global Lung Function 2012 equations. *Eur Respir J* 2012; 40: 1324-1343.

24 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.

25 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.

26 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.

27 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*
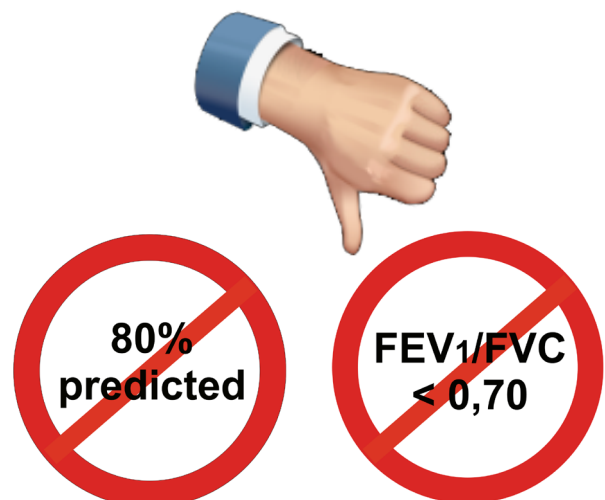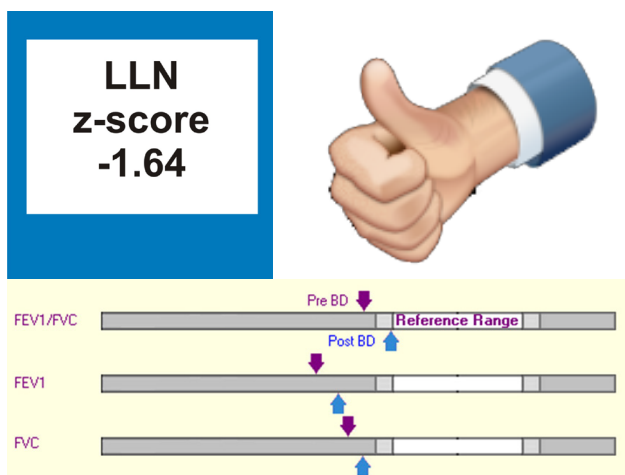
1996; 51: 277-283.

28 Pistelli F, Bottai M, Viegi G, *et al.* Smooth reference equations for slow vital capacity and flow-volume curve indexes. *Am J Respir Crit Care Med* 2000; 161: 899–905. Erratum in: *Am J Respir Crit Care Med* 2001; 164: 1740.

29 Pistelli F, Bottai M, Carrozzi L, *et al*. Reference equations for spirometry from a general population sample in central Italy. *Respir Med* 2007; 101: 814-825.

30 Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with discussion). *Appl Statist* 2005; 54: 507-554.

31 Quanjer PH, Stanojevic S, Stocks J *et al.*, for and on behalf of the Global Lung Initiative. Changes in the FEV1/FVC ratio during childhood and adolescence: an intercontinental study. *Eur Respir J* 2010; 36: 1391-1399.

32 West GB, Brown JH, Enquist BJ. A general model for the origin of allometric scaling laws in biology. *Science* 1997; 276: 122-126.

33 Quanjer PH, Enright PL, Miller MR *et al.* Open Letter. The need to change the method for defining mild airway obstruction. *Eur Respir J* 2011; 37: 720-722.

34 Ekberg-Aronsson M, Pehrsson K, Nilsson JA, Nilsson PM, Löfdahl CG. Mortality in GOLD stages of COPD and its dependence on symptoms of chronic bronchitis. *Respir Res* 2005; 6: 98.

35 Vaz Fragoso CA, Concato J, McAvay G, *et al.* Chronic obstructive pulmonary disease in older persons: a comparison of two spirometric definitions. *Respir Med* 2010; 104: 1189 - 1196.

36 Pedone C, Scarlata S, Sorino C, Forastiere F, Bellia V, Antonelli Incalzi R. Does mild COPD affect prognosis in the elderly? *BMC Pulm Med* 2010; 10: 35.

37 Mannino DM, Doherty DE, Buist AS. Global Initiative on Obstructive Lung Disease (GOLD) classification of lung disease and mortality: findings from the Atherosclerosis Risk in Communities (ARIC) study. *Respir Med* 2006; 100: 115–122.

38 Vaz Fragoso C, Gill T, McAvay G, *et al.* Use of lambda-mu-sigma-derived Z score for evaluating respiratory impairment in middle-aged persons. *Respir Care* 2011; 56: 1771-1777.

39 Bridevaux P-O, Gerbase MW, Probst-Hensch NM, Schindler C, Gaspoz JM, Rochat T. Long-term decline in lung function, utilisation of care and quality of life in modified GOLD stage 1 COPD. *Thorax* 2008; 63: 768 - 774.

40 Mannino DM, Buist AS, Vollmer WM. Chronic obstructive pulmonary disease in the older adult: what defines abnormal lung function? *Thorax* 2007; 62: 37–241.

41 Vaz Fragoso CA, Concato J, McAvay G, *et al*. The ratio of FEV1 to FVC as a basis for establishing chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010; 181: 446 - 451.

42 Borsboom GJJM, van Pelt W, van Houwelingen HC, van Vianen BG, Schouten JP, Quanjer PH. Diurnal variation in lung function in subgroups from two Dutch populations. Consequences for longitudinal analysis. *Am J Respir Crit Care Med* 1999; 159: 1163–1171.

43 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.

44 Knudson RJ, Lebowitz MD, Holberg CJ, *et al.* Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis* 1983; 127: 725–734.

45 Rosenthal M, Bain SH, Cramer D, *et al.* Lung function in white children aged 4–19 years: I – Spirometry. *Thorax* 1993; 48: 794–802.

46 Wang X, Dockery DW, Wypij D, *et al.* Pulmonary function between 6 and 18 years of age. *Pediatr Pulmonol* 1993; 15: 75–88.

47 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.

48 Aaron SD, Dales RE, Cardinal P. How accurate is spirometry at predicting restrictive pulmonary impairment? *Chest* 1999; 115: 869–873.

49 Glady CA, Aaron SD, Lunau ML, *et al.* A spirometry-based algorithm to direct lung function testing in the pulmonary function laboratory. *Chest* 2003; 123: 1939–1946.

50 Swanney MP, Beckert LE, Frampton CM, *et al.* Validity of the American Thoracic Society and other spirometric algorithms using FVC and Forced Expiratory Volume at 6 s for predicting a reduced total lung capacity. *Chest* 2004; 126: 1861–1866.

51 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.

52 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.

53 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.

54 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.

55 Lim LLY, Seubsman S-A, Sleigh 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.

56 Wada K, Tamakoshi K, Tsunekawa T *et al.* Validity of self-reported height and weight in a Japanese workplace population. *Intern J Obesity* 2005; 29: 1093–1099.

57 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. *Eur Respir J* 2012; 40: 190–197.

58 Lum S, Bonner R, Kirkby J, Sonnappa S, Stocks J. S33 Validation of the GLI-2012 multi-ethnic spirometry reference equations in London school children. *Thorax* 2012; 67: A18 (http://thorax.bmj.com/content/67/Suppl_2/A18.2).

59 Hall GL, Thompson BR, Stanojevic S, *et al.* The Global Lung Initiative 2012 reference values reflect contemporary Australasian spirometry. *Respirology* 2012; 17: 1150–1151.

60 Pellegrino R. Viegi G. Brusasco V, *et al.* ATS/ERS Task Force. Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948-968.

61 Quanjer PH, Brazzale DJ, Boros PW, Pretto JJ. Implications of adopting the Global Lungs 2012 all-age reference equations for spirometry. *Eur Respir J* 2013; 32(4): 1046-1054.