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

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 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:

\[
Y = a + b \cdot \text{height} + c \cdot \text{age} + \text{error} \quad (\text{adults})
\]

\[
\log(Y) = a + b \cdot \log(\text{height}) + \text{error} \quad (\text{children})
\]
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. 4 - Relationship between age and FEV1 in 28,690 white, healthy females. About half of the scatter is due to differences in standing height.

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 interpreting 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...
GLI-2012 reference values for spirometry

in scores of publications [10,16,18-23]. In fact, Sobol wrote [19]: “Nowhere else in medicine is such a naive 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 FEV\textsubscript{1} 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], 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:

\[ \log(Y) = a + b \cdot \log(\text{height}) + c \cdot \log(\text{age}) + \text{spline} + \text{error} \]

This approach was adopted by Pistelli et al. [28-29]. However, the statistical package GAMLS [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 FEV\textsubscript{1} 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 FEV\textsubscript{1} 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.

FEV\textsubscript{1}/FVC: a surprise

Analysis of the FEV\textsubscript{1}/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 FEV\textsubscript{1}/FVC ratio in and around the 10 year age range, this could explain the findings. However, no centre had contributed
GLI-2012 reference values for spirometry

Evidence that the findings were not an artefact came from analysis of data from boys and girls from 15 difference centres, comprising different ethnic groups (figure 11, [31]). As the determinants of the FEV\textsubscript{1} and the VC are not the same, it follows that after birth the vital capacity grows proportionally faster than the FEV\textsubscript{1}, and that this pattern is temporarily reversed during the adolescent growth spurt [31].

"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 "normal" test results, whereas in 2.5% they are "too low" and in 2.5% "too high", resulting in 5% false-positive test results. Results of spirometric tests characteristically lead to values for FEV\textsubscript{1} 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 5\textsuperscript{th} centile of a healthy population of non-smokers.

Fig. 11 - Data from 15 centres, comprised of different ethnic groups, all display the same pattern: rapid decline of FEV\textsubscript{1}/FVC ratio until the start of the adolescent growth spurt, then a small increase followed by a decline.

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:

\[
\log(\text{CoV}) = a + b\cdot\log(\text{age}) + \text{spline} + \text{error}
\]

The coefficient of variation for FEV\textsubscript{1} 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 5\textsuperscript{th} 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 FEV\textsubscript{1} 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.

"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 "normal" test results, whereas in 2.5% they are "too low" and in 2.5% "too high", resulting in 5% false-positive test results. Results of spirometric tests characteristically lead to values for FEV\textsubscript{1} 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 5\textsuperscript{th} centile of a healthy population of non-smokers.

Fig. 11 - Data from 15 centres, comprised of different ethnic groups, all display the same pattern: rapid decline of FEV\textsubscript{1}/FVC ratio until the start of the adolescent growth spurt, then a small increase followed by a decline.

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:

\[
\log(\text{CoV}) = a + b\cdot\log(\text{age}) + \text{spline} + \text{error}
\]

The coefficient of variation for FEV\textsubscript{1} 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 5\textsuperscript{th} 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 FEV\textsubscript{1} 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. 12 - Relationship between standard deviation and percentage of data under the curve in the case of a normal distribution.

Fig. 13 - The coefficient of variation (CoV) for FEV\textsubscript{1} in healthy white females varies with age.
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.

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:

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

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>
<thead>
<tr>
<th>Ethnicity</th>
<th>FEV₁ (% of white)</th>
<th>FVC (% of white)</th>
<th>FEV₁/FVC</th>
<th>FEV₁ (% of white)</th>
<th>FVC (% of white)</th>
<th>FEV₁/FVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>African-American</td>
<td>-13.8</td>
<td>-14.4</td>
<td>0.6</td>
<td>-14.7</td>
<td>-15.5</td>
<td>0.8</td>
</tr>
<tr>
<td>North East Asian</td>
<td>-0.7</td>
<td>-2.1</td>
<td>1.1</td>
<td>-2.7</td>
<td>-3.6</td>
<td>0.9</td>
</tr>
<tr>
<td>South East Asian</td>
<td>-13.0</td>
<td>-15.7</td>
<td>2.9</td>
<td>-9.7</td>
<td>-12.3</td>
<td>2.8</td>
</tr>
<tr>
<td>Mixed/other</td>
<td>-6.8</td>
<td>-7.9</td>
<td>1.1</td>
<td>-6.8</td>
<td>-7.9</td>
<td>1.1</td>
</tr>
</tbody>
</table>
that it is necessary to adjust the model for the coefficient of variation, shown earlier, as follows:

$$\log(\text{CoV}) = a + b \cdot \log(\text{age}) + d \cdot \text{Ethn} + \text{spline} + \text{error}$$

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...
interpreting the findings in the wink of an eye (fig. 20 and 21).

**Comparison of predicted values**

Paediatricians in the Netherlands rely almost 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 according to ECCS/ERS [10] and NHANES [24] differ 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 equations only take age into account. Predicted values 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. 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.
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 range of diagnoses from two hospitals in Australia and one in Poland (fig. 25) disclosed the following trend (fig. 26). There was 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 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).

“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.

<table>
<thead>
<tr>
<th>Author</th>
<th>Boys (n = 2492)</th>
<th>Girls (n = 2072)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hankinson</td>
<td>17.8%</td>
<td>14.3%</td>
</tr>
<tr>
<td>Knudson</td>
<td>21.0%</td>
<td>10.5%</td>
</tr>
<tr>
<td>Quanjer GLI-2012</td>
<td>15.0%</td>
<td>14.0%</td>
</tr>
<tr>
<td>Wang</td>
<td>21.6%</td>
<td>16.8%</td>
</tr>
<tr>
<td>Zapletal</td>
<td>23.1%</td>
<td>10.9%</td>
</tr>
</tbody>
</table>

Fig. 22 - Comparison of predicted FEV₁ and FVC in healthy boys and girls according to GLI-2012 [23], Zapletal [43], Stanoevic [16], Polgar [12], Quanjer [13], Hankinson [24], Knudson [44], Rosenthal [45] and Wang [46].

Fig. 23 - Comparison of predicted FEV₁/FVC ratio in boys and girls according to GLI-2012 [23], Hankinson [24] and ECCS/ERS [10].
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
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, where $k \sim 2.2$. 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...
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 FEV₁ and FVC.

<table>
<thead>
<tr>
<th>Age (yr) (rounded off)</th>
<th>Males</th>
<th>Females</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FEV₁</td>
<td>FVC</td>
<td>FEV₁</td>
<td>FVC</td>
</tr>
<tr>
<td>3 vs 3.75</td>
<td>-2.8</td>
<td>-3.4</td>
<td>-2.9</td>
<td>-3.6</td>
</tr>
<tr>
<td>10 vs 10.75</td>
<td>-1.3</td>
<td>-1.4</td>
<td>-2.6</td>
<td>-2.7</td>
</tr>
<tr>
<td>15 vs 15.75</td>
<td>-3.4</td>
<td>-2.9</td>
<td>-3.4</td>
<td>-2.9</td>
</tr>
<tr>
<td>50 vs 50.75</td>
<td>+0.4</td>
<td>+0.4</td>
<td>+0.6</td>
<td>+0.7</td>
</tr>
<tr>
<td>85 vs 85.75</td>
<td>+0.7</td>
<td>+0.5</td>
<td>+0.9</td>
<td>+1.0</td>
</tr>
</tbody>
</table>

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](https://www.gluco.org/software).

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](https://www.gluco.org/software).

**Flows**

There are recurrent questions why predicted values for instantaneous flows, such as FEF50, have not been included in the GLI-2012 set. These flows have never been shown to have added value over and above FEV₁ and VC. These flows are often considered to be sensitive indices of "small airways 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 FEF75% and FEF25-75%.

**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 O₂ 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 to ECSC: “treat the patient, not the data”.

Acknowledgements

Figure 22: Modified and reproduced with permission of the European Respiratory Society. Eur Respir J July 2012 40:190-197; published ahead of print December 19, 2011, doi:10.1183/09031936.00161011
Figure 25 and 29: Modified and reproduced with permission of the European Respiratory Society. Eur Respir J 2013; in press; doi: 10.1183/09031936.00195512.

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.
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.