



# Choosing and using lung function prediction equations

Martin R. Miller

**Affiliation:** Institute of Occupational and Environmental Medicine, University of Birmingham, Birmingham, UK.

**Correspondence:** Martin R. Miller, Institute of Occupational and Environmental Medicine, University of Birmingham, Birmingham, B15 2TT, UK. E-mail: m.r.miller@bham.ac.uk



@ERSpublications

**The GLI-2012 reference equations give the best fit for patients undergoing spirometry lung function testing** <http://ow.ly/NikG304sh0n>

In a survey of lung function laboratories some years ago, a surprisingly large number of laboratories were found not to know exactly which equations were being used to help them make judgements about their patients' results [1]. Many manufacturers' equipment includes a choice of reference values and the default selection may end up being used without further or due consideration. When choosing which equations to use, laboratory directors and clinicians may worry that their patients might be assessed against an inappropriate standard, thus introducing potential errors in their decision making. The European Respiratory Society has sponsored the Global Lung Function Initiative (GLI) in obtaining coherent reference equations for spirometry appropriate for all ages [2] and advocates these being the standard applied for spirometry within Europe. In this issue of the *European Respiratory Journal* (ERJ), two groups of European researchers have looked at their local populations to see if the GLI-2012 equations were a satisfactory fit and so can be safely used for clinical decision making [3, 4].

HULO *et al.* [3] looked at the data of 1971 people aged 40–65 years from an urban population who were symptom free and were part of a study of 3276 subjects who were sampled to determine the prevalence of airflow obstruction in the general population. The mean $\pm$ SD z-scores for forced expiratory volume in 1 s (FEV<sub>1</sub>) and forced vital capacity (FVC) by GLI-2012 were close to 0.0 $\pm$ 1.0, as expected, and were better than the scores found when using the European Community for Steel and Coal (ECSC) equations [5]. Their subjects' FEV<sub>1</sub>/FVC results were slightly low relative to GLI-2012, with mean z-scores of –0.32 and –0.40 for men and women, respectively, but they were not able to account for smoking in their population, which may have influenced this. They also looked at how the two equations fared at identifying abnormalities in another 1072 subjects with respiratory disease and concluded that the GLI-2012 equations were better than the ECSC equations in this respect.

LANGHAMMER *et al.* [4] used data from 11 496 Norwegian subjects across an age range of 12–90 years who were healthy never-smokers without symptoms. They found that the median z-scores for FEV<sub>1</sub>, FVC and FEV<sub>1</sub>/FVC were essentially zero in those aged <20 years, with female adults tending to have very slightly better FEV<sub>1</sub> and FVC and lower FEV<sub>1</sub>/FVC than the values predicted by GLI-2012 (median z-scores 0.15, 0.25 and –0.17, respectively). They concluded that GLI-2012 was a better fit for their data than either the ECSC or an alternative Norwegian prediction equation, and recommended GLI-2012 for Norwegian subjects. They also applied GLI-2012 to a larger population including symptomatic subjects (n=50 580) and concluded that GLI-2012 was better than FEV<sub>1</sub>/FVC <0.7 at identifying clinically relevant airflow obstruction.

Both these studies found an acceptable fit for their data when using GLI-2012, with some small differences noted. In this issue of the ERJ, QUANJER and STANOJEVIC [6] have also outlined some reasons as to why local populations may appear to give an imperfect fit for the GLI-2012 predictions. The strength of GLI is that

---

Received: Sept 01 2016 | Accepted: Sept 02 2016

Conflict of interest: None declared.

Copyright ©ERS 2016

for each sex the equations are based on the data of >35 000 subjects. Small sample sizes from other populations tested against this may appear to be different solely due to sampling issues. European and North American populations are well represented by GLI and so the need for further large sample validation of these populations is not a pressing issue. However, QUANJER and STANOJEVIC [6] indicate that the large data resource behind the GLI project needs additional data of satisfactory quality from under-represented populations, for example from India, Africa and Latin America to name just three, in order to be sure that these populations are better able to be assessed in the future. A further difficulty is obtaining adequate sampling of representative people of advanced age with normal lungs; with the ever-increasing numbers of people surviving to advanced age, this continues to be an area of uncertainty that needs addressing [7].

All of the aforementioned points give reassurance that with GLI-2012 the right reference standard is being applied. How patients' results are then used in relation to their predicted values is where there is potential scope for more error than that related to the choice of equations. QUANJER and STANOJEVIC [6] reiterate that using "percentage of predicted" for determining if a result is unusual is completely unacceptable, because the method retains age, sex and size bias. The lower limit of normal (LLN) should be used to decide if a result is unusual and the LLN has been taken to be the lower fifth percentile ( $z$ -score  $< -1.645$ ) for subjects referred for lung function testing [8]. This means that if entirely normal healthy subjects are tested, around 5% of them will be found to be below this limit and yet they do not have anything wrong with their lung function. This implicit false positive rate of 5% may seem high but is deemed acceptable since lung function tests are usually only undertaken on subjects where the *a priori* probability of an abnormality being present is already high due to the presence of symptoms, signs, exposures or events that have led to the referral for testing. Thus, the LLN of the fifth percentile is acceptable and trades some increased sensitivity for a possible lower specificity. If a test is below the LLN this indicates that the result is potentially unusual for someone with normal lungs and the clinician must interpret this in conjunction with the clinical indications that led to the referral for testing. However, if lung function testing is to be undertaken on a population not specially selected as having an increased risk of abnormality, for example random sampling or screening, then the lower 2.5th percentile ( $z$ -score  $< -1.96$ ) is more appropriate.

Correction for ethnicity remains potentially problematic but can be undertaken using the patient's self-reported ethnicity and a specific or amalgamated correction factor in GLI-2012 [2]. In France, the registration of ethnicity is not allowed by law, on the basis that it might infringe an individual's right to equality of treatment. However, not taking known variations in lung function due to ethnicity into account may in fact lead to inequality in treatment because individuals from certain ethnic groups may have their lung function incorrectly deemed abnormal (or excessively low) because they were being judged against an inappropriately high standard. In these instances, clinicians must be aware of this possibility before making clinical decisions based on these results alone. Another determinant of lung size and function is the biological sex of the subject and so for transgender patients this should be used in the prediction equation rather than the stated gender. In the rare androgen insensitivity syndrome, the lungs and subject are biologically female although the subject is chromosomally XY. Age and height are the other determinants used to predict lung function and it is agreed that age must be expressed to at least one decimal place and height to the nearest centimetre. Truncating age and a 1-cm height bias have been estimated to cause up to 8% and 1% errors, respectively, in FEV<sub>1</sub> and FVC predictions [9].

Before acting upon an abnormal clinical result, the prudent clinician must first be certain that the possibility of errors within the result and its interpretation have been minimised. For patients undergoing spirometry, this will be best achieved by the use of the GLI-2012 equations, and GLI equations for transfer factor of the lung for carbon monoxide are anticipated in the near future.

## References

- 1 Dowson LJ, Mushtaq M, Watts T, *et al.* A re-audit of pulmonary function laboratories in the West Midlands. *Respir Med* 1998; 92: 1155–1162.
- 2 Quanjer PH, Stanojevic S, Cole TJ, *et al.* Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012; 40: 1324–1343.
- 3 Hulo S, de Broucker V, Giovannelli J, *et al.* Global Lung Function Initiative reference equations better describe a middle-aged, healthy French population than the European Community for Steel and Coal values. *Eur Respir J* 2016; 48: 1779–1781.
- 4 Langhammer A, Johannessen A, Holmen TL, *et al.* Global Lung Function Initiative 2012 reference equations for spirometry in the Norwegian population. *Eur Respir J* 2016; 48: 1602–1611.
- 5 Quanjer PH, Tammeling GJ, Cotes JE, *et al.* Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J* 1993; 6: Suppl. 16, 5–40.
- 6 Quanjer PH, Stanojevic S. Do the Global Lung Function Initiative 2012 equations fit my population? *Eur Respir J* 2016; 48: 1782–1785.

- 7 Miller MR, Thinggaard M, Christensen K, *et al.* Best lung function equations for the very elderly selected by survival analysis. *Eur Respir J* 2014; 43: 1338–1346.
- 8 Pellegrino R, Viegi G, Brusasco V, *et al.* Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.
- 9 Quanjer PH, Hall GL, Stanojevic S, *et al.* Age- and height-based prediction bias in spirometry reference equations. *Eur Respir J* 2012; 40: 190–197.