



Global Lung Function Initiative

Lung Function in Growth and Aging a united worldwide approach

Introduction

This software is designed for PC ONLY and is provided 'as is', without any warranty. The output has been extensively tested against the output from GAMLSS in R, and found to be accurate.

The software can be used to:

- 1 Add predicted values and z-scores to an existing dataset.
- 2 Check (by IT engineers) whether the implementation in software of the GLI equations and lookup file leads to correct results.
- 3 Validate application of the GLI equations to one's own dataset.

What the software does:

- 1 Reads a tab-separated data file (extension .txt)
- 2 Calculates predicted values, their lower and upper limits of normal (5th and 95th centiles), percent of predicted, and z-scores for FEV₁, FVC, FEV₁/FVC, FEV_{0.75}, FEV_{0.75}/FVC, FEF₇₅, FEF_{25-75%}.
- 3 Exports the results to a tab-separated (.txt) file for manipulation in a spreadsheet or other software.

Preparing your tab-separated text file

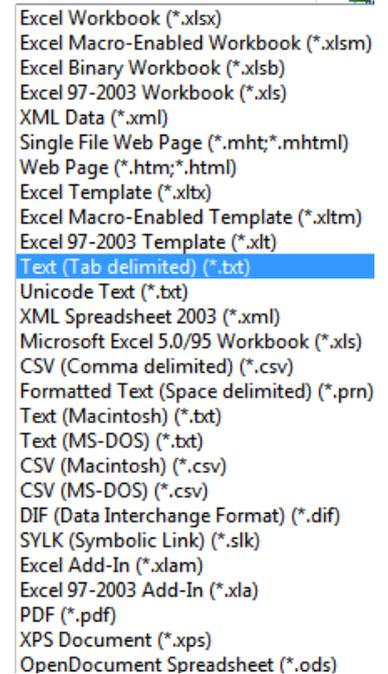
IT IS IMPORTANT TO FOLLOW THE INSTRUCTIONS EXACTLY, TO AVOID ERRORS

- 1 Open your data in EXCEL. The maximum number of records is limited to 60,000.
- 2 Ensure you have columns for ALL the variables listed below. If you have not measured a variable, include that column but with all values set to 9999.
- 3 Ensure the columns are in the same order as listed below.

Variables	Units	Example	Limits	Valid age range (years)
ID	integer	411		
Sex	integer	1(male), 2(female)	1 – 2	
Age	years	45.57	3.0 – 95.0	
Height	cm	143.7	50 – 250	
Ethnicity	integer	1(white), 2(black), 3(NE Asia*), 4(SE Asia*) 5(Other/mixed§)	1 – 5	
FEV ₁	litres	3.456	0.2 – 9.0	3 – 95
FVC	litres	4.035	0.3 – 11.0	3 – 95
FEV ₁ /FVC		0.85	0.2 – 1.0	3 – 95
FEV _{0.75}	litres	2.862	0.2 – 6.0	3 – 7
FEV _{0.75} /FVC		0.76	0.2 – 1.0	3 – 7
FEF _{25-75%}	L/s	3.786	0.1 – 11.0	3 – 95
FEF ₇₅	L/s	2.305	0.02 – 8.0	3 – 90

* The dividing line between Korea and China (North) and South East Asia is the Huaihe River and Qinling Mountains. § Other group or mixed ethnic descent.

- 4 Ensure the units match and that the data are within the limits specified. If data are outside the limits the software usually issues a warning and exits, but in some cases the user is allowed to continue if certain that the offending data are correct.
- 5 Code all missing values as 9999.
- 6 Ensure that sex, age, height and ethnic group are not missing, and that height and age are entered to at least 1 decimal place to maximise accuracy (Quanjer *et al.*, *Eur Respir J* 2012; 40: 190–197).
- 7 Save the file as a tab-separated text (.txt) file.



Period or comma as decimal separator

Depending on the language settings of your system, the decimal separator may be a comma (e.g. France, Germany) or a period (English speaking countries). The examples shown, and TestFile.txt (see later) all use the decimal period. The software converts decimal periods to commas, and *vice versa*, as required.

Starting the software

Upon opening GLI2012_DataConversion.EXE, 'Database conversion' is checked by default in 'Select Output' (Figure 1). Click 'File' and select the file that you want to convert. The content of the file and a button 'Calculate' are displayed (Figure 2). Click on the button and save (see under 'File') the resulting file (with extension .txt) when prompted.

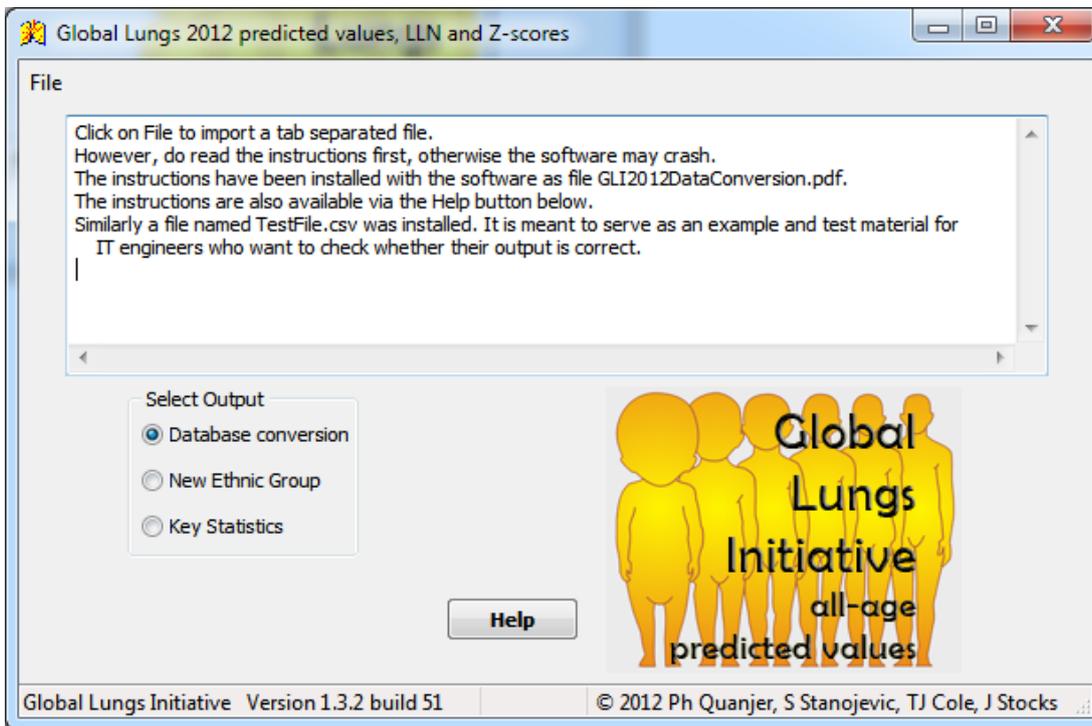


Figure 1 – Opening screen of GLI2012_DataConversion.EXE.

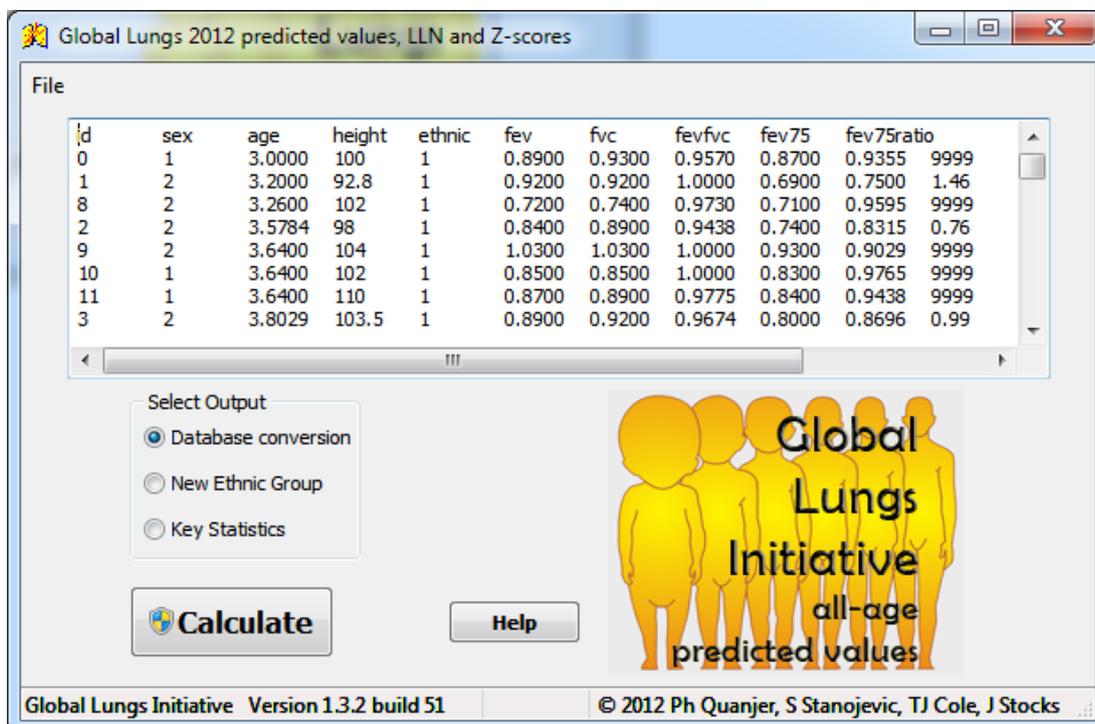


Figure 2 – Screen after importing a file for conversion.

If the software crashes and you are absolutely certain that your dataset contains no errors, please note carefully what steps led to the crash and contact pquanjer@lungfunction.org.

Countries which contributed data

Caucasians	Europe, Israel, Australia, USA, Canada, Mexican Americans, Brazil, Chile, Mexico, Uruguay, Venezuela, Algeria, Tunisia
Blacks	African Americans
South East Asians	Thailand, Taiwan and China (including Hong Kong) south of the Huaihe River and Qinling Mountains
North East Asians	Korea and China north of the Huaihe River and Qinling Mountains

Test file for manufacturers – IT engineers

Manufacturers can test whether their implementation of equations in software leads to correct results. To that end they can use the same age, height *etc.* as in TestFile.txt and compare their output to that in Validate.txt. Anyone can create one's own test file and process this with the GLI2012_DataConversion.EXE software; the output file contains the original data, predicted values, lower and upper limits of normal (5th and 95th centiles), percent of predicted, and z-scores to 4 decimal accuracy. TestFile.txt and the GLI2012_DataConversion software are included in the installation file, available from www.lungfunction.org/files/InstallGLI2012_DataConversion.EXE.

Deriving coefficients for additional ethnic groups

The GLI2012_DataConversion.EXE software can be used to derive coefficients for additional ethnic groups from the GLI 2012 equations. A prerequisite is that a sufficiently large group of healthy nonsmokers, studied according to current international standards and appropriate quality control, is available. A minimum group size is 150 males and 150 females. Please note that samples of this size may differ by up to + 0.4 z-scores from the true population mean, and by up to +0.3 z-scores even if sample size exceeds 1,000 subjects (*Eur Respir J* 2011; 37: 658–664).

In order to calculate coefficients for a new group (*e.g.* Inuit, highlanders in Bolivia, Australian aboriginals) to the set of equations, proceed as follows.

- 1 Format the group as indicated in the introduction (see table page 1). The only difference is that the ethnicity must be recorded as 1 (Caucasian), in order to calculate how much the new group differs

from the Caucasian group.

- 2 Start the GLI2012_DataConversion software, select "New Ethnic Group", open your file, press Calculate, and export the resulting file (with .txt extension).

The following output is produced for FEV₁, FVC, FEV₁/FVC, FEF_{25-75%}, and FEF₇₅:

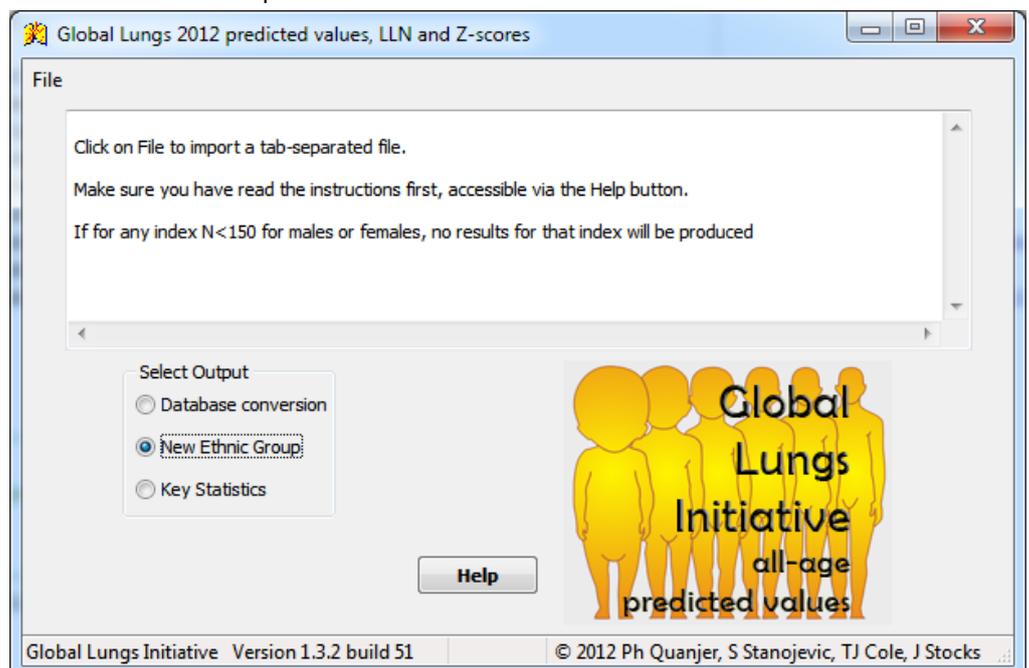
- Average z-score for males and females: avgFEVM and avgFEVF are the mean z-scores for FEV₁ in males and females, respectively.
- The standard deviation (SD) of z-scores for the various indices. Thus sdFVCM and sdFVCF represent the SD for the z-scores for FVC in males and females, respectively.
- The coefficients for M and S for your new group. Thus Mratiogroup and Srationgroup represent the coefficients for FEV₁/FVC to be applied to your new group. These coefficients should be applied to both males and females.

Note that output will only be produced for FEV_{0.75} and FEV_{0.75}/FVC in Caucasians, as GLI 2012 predicted values for these indices are not available for other groups. Even so, your data file should be formatted as delineated in the introduction, where these indices may be coded as missing (code value as 9999).

As shown in the GLI 2012 report, there are essentially no inter-group differences in the FEV₁/FVC ratio. If you find a large difference in this ratio between your new group and predicted values for Caucasians, you should be concerned about your measurements.

Implementing additional coefficients into commercial lung function devices

It will be possible for software manufacturers to implement the facility to enter additional coefficients for calculating predicted values for new ethnic groups, provided written agreement has been obtained from the



GLI analytical team. This will be subject to the following conditions:

- Individuals or groups requiring inclusion of such updates need to provide the GLI team with the following:
- 1 Written confirmation that data were collected with full ethical approval and with agreement for publication.
 - 2 Confirmation that data were collected using international standards for technical acceptability, valid at time of measurement.
 - 3 Agreement to share with the Global Lung Function Initiative the data from which coefficients have been derived, so that it can be incorporated into any future dataset or combined with other similar ethnic groups.
 - 4 Submission of individual spirometry results, and relevant background details (age, sex, height, ethnic group, etc.) in standardised anonymised format.
 - 5 Agreement that details of the additional coefficients should be open access/shared with all manufacturers.

Validating datasets

If data have collected from at least 150 healthy non-smoking males and at least 150 healthy non-smoking females, it is possible to check how well the data agree with GLI2012 predicted values.

- A tab-separated file should be prepared as described on page 1.
- Select 'Key Statistics'
- Click 'File' and open the txt file
- When the button 'Calculate' appears, click the button

- Click file again, and save the file with extension .txt
- A sample of the output is presented below.

Z-scores FEV ₁	males	females
Mean	-0.5173	-0.5800
SD	1.4123	0.9343
Minimum	-4.7594	-5.2559
Maximum	4.0994	3.4798
% z-score < -1.6445	9.8	12.0
% z-score > 1.6445	1.0	1.1

Ideally the average z-score (here only shown for FEV₁) should be 0, the standard deviation 1, and about 5% of observations should be below the 5th and above the 95th centile. In large datasets (N > 1,000) differences of about 0.3 z-scores are observed, in small datasets (N = 150-1,000), differences may be up to 0.4 z-scores (*Eur Respir J* 2011; 37: 658–664). In this example, based on a large dataset, observed values are systematically below target. Individuals were classified as Caucasians, which was not their true ethnicity. The findings show that in this ethnic group observed values are 0.5 to 0.6 z-scores below those for Caucasians. The SD for FEV₁ in males is too large; this should prompt a search for causes of an unusually large variability.

Analytical Team of Global Lung Function Initiative
 Philip H Quanjer
 Sanja Stanojevic
 Tim J Cole
 Janet Stocks
 4 September 2012

