Supplementary File 2: Statistical Models Fitted

(ERJ manuscript number ERJ-02129-2019.R2)

Following Quanjer et al. [2] and Cole et al. [13], we applied the GAMLSS methodology to fit flexible statistical models to each of the outcomes FEV_1, FVC and FEV_1/FVC; with age and/or height included as predictors of up to three distribution parameters: location (μ), scale (σ) and skewness (ν). Separate models were fitted within male and female groups (i.e. we used a completely stratified analysis).

As suggested by Quanjer et al. [2]and Cole et al. [13], models were chosen according to minimising the Schwarz Bayesian Criterion (SBC), to achieve the most parsimonious final model and protect against overfitting. We selected models according to (i) the assumed outcome distribution, (ii) the linear predictors of the distribution parameters, (iii) the necessity of spline functions (flexible functions which can model curved and non-linear relationships), and (iv) the necessity of interaction terms. Regarding the assumed outcome distribution, we focussed on continuous distribution models with two parameters each (especially Gamma, Inverse Gaussian, and Normal models); plus the Box-Cox-Cole-Green (BCCG) distribution with three distributional parameters, which has previously been used to model the outcome variables in different populations [2, 13].

Regarding linear predictors of the distribution parameters, we started with models including age and height as predictors of each parameter; but we frequently encountered convergence problems so invariably settled for a reduced number of predictors. For spline functions, we considered penalised beta splines (P-splines) which have advantages in efficiency and reliability over alternative spline functions [14], although in fact none were found to reduce the SBC compared to linear terms.

Specifically, we fitted the following models to the data:

• FEV_1 in females was modelled as Inverse Gaussian, with log link functions for the parameters μ and σ, including age and height as predictors for μ, and age for σ.

• FVC in females was modelled as Gamma, with log link functions for the parameters μ and σ, including age and height as predictors for μ, and age for σ.

• FEV_1/FVC in females was modelled as Normal, with log link functions for the parameters μ and σ, including age as predictor for μ, and none for σ.

Supplementary File 2: Statistical Models Fitted

(ERJ manuscript number ERJ-02129-2019.R2)

- FEV$_1$ in males was modelled as Normal, including age and height as predictors for μ, and none for σ (i.e. σ was constant). This was equivalent to fitting a linear regression model with age and height.

- FVC in males was modelled as Normal, including age and height as predictors for μ, and none for σ (i.e. σ was constant). This was equivalent to fitting a linear regression model with age and height.

- FEV$_1$/FVC in males was modelled as Normal, with identity link function for the parameter μ, including age as predictor for μ, and none for σ. This was equivalent to fitting a linear regression model with age.