APPENDIX D

REFERENCE AND PREDICTIVE VALUES FOR PEAK EXPIRATORY FLOW RATE (PEFR)

Lung function is related to physical characteristics such as age and height. In order to assess the Peak Expiratory Flow Rate (PEFR) measurement it is necessary to establish reference values for PEFR for the South African population. These reference values, based on a healthy asymptomatic sample can then be used to standardise the PEFR of every individual in the sample for his or her physical characteristics.

A healthy sub-sample of the SADHS adult sample was created by excluding individuals with the following reported health and lung problems or status:
1) smokers
2) asthma
3) chronic bronchitis
4) reported asthma
5) reported TB
6) reported emphysema/bronchitis
7) reported lung cancer
8) pregnant women.

After these exclusions the data of the remaining sample was checked for outliers and inconsistencies in the data. The final sample size for the PEFR reference values is given in Table D1.

<table>
<thead>
<tr>
<th>Table D1. Sample sizes of the PEFR reference sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
</tr>
<tr>
<td>Total SADHS</td>
</tr>
<tr>
<td>Total PEFR</td>
</tr>
<tr>
<td>Total PEFR standardised</td>
</tr>
<tr>
<td>Total healthy sub-sample</td>
</tr>
</tbody>
</table>

PEFR was not measured in about 300 adults and another 100 adults had missing information on the variables used for the standardisation of PEFR.

Estimation of Reference Values

The modeling of PEFR in the healthy sub-sample consists of two components:
i) the mean predicted value for a male or female of a specific age, height and weight
ii) the mean predicted standard deviation in the population at the covariate values of (i).

Linear regression models were used for both of the components. The regression was done using only the basic characteristics of each subject: age in years, weight in kg, height in meters and sex. These are shown in Table D2 for men and women.
Table D2. Descriptive statistics for the healthy sub-sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>31</td>
<td>36</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>64</td>
<td>67</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.67</td>
<td>1.60</td>
</tr>
<tr>
<td>PEFR (liters/min)</td>
<td>370</td>
<td>306</td>
</tr>
</tbody>
</table>

In the case of the mean predicted value, the dependent variable PEFR was transformed by taking the natural logarithm of PEFR as a variance stabilizing transformation.

In the case of the mean predicted standard deviation, the absolute value of the residuals from the log(PEFR) model was used as the dependent variable (Altman, 1993).

Regression Models

Both models consist of high order polynomials in the predictor variables and have interaction terms especially for sex to model the regression function for males and females. The higher order polynomials used, reflect the non-linear relationship of PEFR over especially age and weight ranges. All three the predictor variables (age, height and weight) were evaluated for the significance of fourth order polynomials and interactions. Since the models were developed for predictive purposes only the significant terms were included in the final model.

The inverse of height was used as a predictor which corresponds to the adjustment used for height in the calculation of the body mass index.

The estimated parameters of the two regression models are given in Table D3.

Table D3. Estimation equations

Regression model for log (pefr)

\[
\text{pred\_log(pefr)} = 4.961423416 + \\
-0.021814972 \times \text{age} + \\
0.000831883 \times \text{age}^2 + \\
-0.00013905 \times \text{age}^3 + \\
0.000000072 \times \text{age}^4 + \\
-0.227671804 \times \text{sex} + \\
0.048494426 \times \text{age}\times \text{sex} + \\
-0.001711697 \times \text{age}^2 \times \text{sex} + \\
0.000024669 \times \text{age}^3 \times \text{sex} + \\
-0.000000126 \times \text{age}^4 \times \text{sex} + \\
0.405514263 \times \text{ilht} + \\
0.037593478 \times \text{wht} + \\
-0.000268379 \times \text{wht}^2 + \\
0.000000906 \times \text{wht}^3 + \\
-0.019598017 \times \text{ilht}\times \text{wht} + \\
0.001108633 \times \text{wht}\times \text{sex} + \\
-0.269653218 \times \text{ilht}\times \text{sex}
\]

R-square = .214 Root MSE = .2541 Mean log (PEFR) = 5.75025
Regression model for the standard deviation of log (pefr)

\[
\text{sd\_log(pEFR)} = \sqrt{3.141592654/2} \times \\
(0.0468321524 + \\
0.0121021061 \times \text{age} + \\
-0.0003260880 \times \text{age}^2 + \\
0.0000003264 \times \text{age}^3 + \\
-0.0000000085 \times \text{age}^4 + \\
-1.984145411 \times \text{sex} + \\
0.0225272293 \times \text{age} \times \text{sex} + \\
-0.0009063496 \times \text{age}^2 \times \text{sex} + \\
0.0000143916 \times \text{age}^3 \times \text{sex} + \\
-0.000000778 \times \text{age}^4 \times \text{sex} + \\
0.0000014714 \times \text{wht} + \\
0.0006113482 \times \text{wht} \times \text{sex})
\]

\[R\text{-square}=0.0117 \quad \text{Root MSE}=0.16033 \quad \text{Mean of absolute residuals}=0.19608\]

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>whit:</td>
<td>weight in kg</td>
</tr>
<tr>
<td>lHt:</td>
<td>1/height (inverse of height in meters)</td>
</tr>
<tr>
<td>sex:</td>
<td>gender indicator with male=1, female=0</td>
</tr>
<tr>
<td>sqrt:</td>
<td>square root</td>
</tr>
<tr>
<td>+:</td>
<td>addition</td>
</tr>
<tr>
<td>*:</td>
<td>multiply</td>
</tr>
<tr>
<td>**:</td>
<td>to the power</td>
</tr>
</tbody>
</table>

**Standardisation**

The standardised values or Z-scores of PEFR is then obtained from the following equation:

\[z_{pefr} = (\log(PEFR) - \text{pred\_log(pEFR)}) / \text{sd\_log(pEFR)}\]

Where a PEFR was more than 2 standard deviations below the predicted value, it was considered abnormal and this cut-off was used in the total adult sample.

**Graphical Displays**

The following two figures display the scatterplot of PEFR versus age by sex. Superimposed on each plot as a smoothed curve which estimates the mean predicted PEFR value across the age range only, based on the healthy sub-sample.
Figure 1: Peakflow versus age in Males from healthy sub-sample

Figure 2: Peakflow versus age in Females from healthy sub-sample
**PEFR Measurements**

The following points about the PEFR measurement should be noted as background to the reference values that have been modeled.

1) The Peak Expiratory Flow Rate meter is a very crude device and the majority of the measurements were rounded to the nearest 50 liters/minute by the field workers.

2) The fieldworkers were required to write a referral note for the participant if he or she could not reach the reference value of 200 liters/minute. This possibly introduced an upward bias in the recordings to the level of 200 and 250 liters/minute or resulted in differential motivation by the fieldworkers so as to avoid this interaction with the respondent. Since the average PEFR for females is 306 l/min (SD=83) compared to 370 l/min (SD=108) for males, the regression models for women will be more affected. By combining the female and male data for the estimation of the reference values, this bias will have been reduced.

3) The maximum reading of the meter is 800 l/min. This censoring was not taken into account in the modeling since it involved a small number of participants.

The nonlinear relationship of PEFR against age, the threshold bias at 200 l/min, the censoring at 800 l/min and the heterogeneous variance across age is evident from both of the figures 1 and 2.