




Normal spirometry predictive values for the Western Indian adult population

Dhiraj Agarwal¹, Richard A. Parker ², Hilary Pinnock³, Sudipto Roy¹, Deesha Ghorpade⁴, Sundeep Salvi⁴, Parag Khatavkar⁵ and Sanjay Juvekar¹ for the RESPIRE collaboration⁶

Affiliations: ¹Vadu Rural Health Program, KEM Hospital Research Centre, Pune, India. ²Edinburgh Clinical Trials Unit, Usher Institute, The University of Edinburgh, Edinburgh, UK. ³NIHR Global Health Research Unit on Respiratory Health (RESPIRE), Usher Institute, The University of Edinburgh, Edinburgh, UK. ⁴Chest Research Foundation, Pune, India. ⁵KEM Hospital, Pune, India. ⁶A list of members of the RESPIRE collaboration can be found in the acknowledgement section.

Correspondence: Sanjay Juvekar, Vadu Rural Health Program, KEM Hospital Research Centre, Pune, India. E-mail: sanjay.juvekar@kemhrcvadu.org



@ERSpublications

The Western Indian adult population appears to have lower lung volumes compared to the Euro-American population. Use of GLI normal values may result in overdiagnosis of respiratory disease and locally derived equations should be used in clinical practice. <https://bit.ly/3aMAN4s>

Cite this article as: Agarwal D, Parker RA, Pinnock H, *et al.* Normal spirometry predictive values for the Western Indian adult population. *Eur Respir J* 2020; 56: 1902129 [<https://doi.org/10.1183/13993003.02129-2019>].

ABSTRACT Interpretation of spirometry involves comparing lung function parameters with predicted values to determine the presence/severity of the disease. The Global Lung Function Initiative (GLI) derived reference equations for healthy individuals aged 3–95 years from multiple populations but highlighted India as a “particular group” for whom further data are needed. We aimed to derive predictive equations for spirometry in a rural Western Indian adult population.

We used spirometry data previously collected (2008–2012) from 1258 healthy adults (aged 18 years and over) by the Vadu Health and Demographic Surveillance System. We constructed sex-stratified prediction equations for forced expiratory volume in 1 s (FEV₁), forced vital capacity (FVC), and FEV₁/FVC using the Generalised Additive Model for Location, Scale and Shape (GAMLSS) method to derive the best fitting model of each outcome as a function of age and height.

When compared with GLI Ethnicity Codes 1 (White Caucasian) and 5 (Other/Mixed), the Western Indian adult population appears to have lower lung volumes on average, though the FEV₁/FVC ratio is comparable. Both age and height were predictive of mean FEV₁ and FVC; and for females, the variability of response was also dependent on age. FEV₁/FVC appears to have a very strong age effect, highlighting the limitations of using a fixed 0.7 cut-off value.

The use of GLI normal values may result in overdiagnosis of lung disease in this population. We recommend that the values and equations generated from this study should be used by physicians in their routine practice for diagnosing disease and its severity in adults from the Western Indian population.

This article has supplementary material available from erj.ersjournals.com

Received: 27 Aug 2019 | Accepted after revision: 24 April 2020

Copyright ©ERS 2020

Background

Spirometry is the gold standard for accurate and repeatable measurement of lung function. It is central to detecting obstructive airway diseases (OADs) such as asthma and chronic obstructive pulmonary disease (COPD) as well as suspecting restrictive lung diseases [1]. Interpretation of spirometry involves assessing lung function parameters such as forced expiratory volume in 1 s (FEV_1), forced vital capacity (FVC) and FEV_1/FVC in comparison with reference predicted values and the absolute values of FEV_1/FVC to determine the presence and severity of OADs. Reference predicted values are typically calculated from thousands of healthy individuals and vary according to not only their age, sex and height, but also their ethnicity [2].

The Global Lung Function Initiative (GLI), reports normative reference values, derived from over 160 000 data points in combined datasets from 72 centres in 33 countries [2]. In this, reference equations were derived for healthy individuals aged 3–95 years principally for European-Americans ($n=57\,395$), African-Americans ($n=3545$), North ($n=4992$) and South East Asians populations ($n=8255$). For individuals not represented by these four groups, or of mixed ethnic origins, a composite equation taken as the average of the above equations has been provided to facilitate interpretation until a more appropriate solution is developed. India is highlighted as a “particular’ group” in whom further data are needed as the limited existing literature suggests the predicted values may be “at least as low as those found in subjects of African origins” [2].

In the absence of predictive values specific to the Indian population, physicians and researchers depend on estimates for non-Indian populations such as GLI [2] or European Community Respiratory Health Survey (ECRHS) [3, 4] with a correction factor to interpret spirometry. This may lead to misclassification of disease in the Indian population if people of Indian origin are different in terms of ethnicity and body habitus compared to other populations. To overcome this, there is a need to develop reference predictive values specific to the Indian population as well as to geographically diverse sub-populations within India (see figure 1 for an overview). Previous studies that have tried to do this have either been based on very small numbers [5–7], limited age range [8], or populations from northern or eastern India [8, 9]. In this study, we aimed to develop reference equations for lung function parameters in the Western Indian adult population.

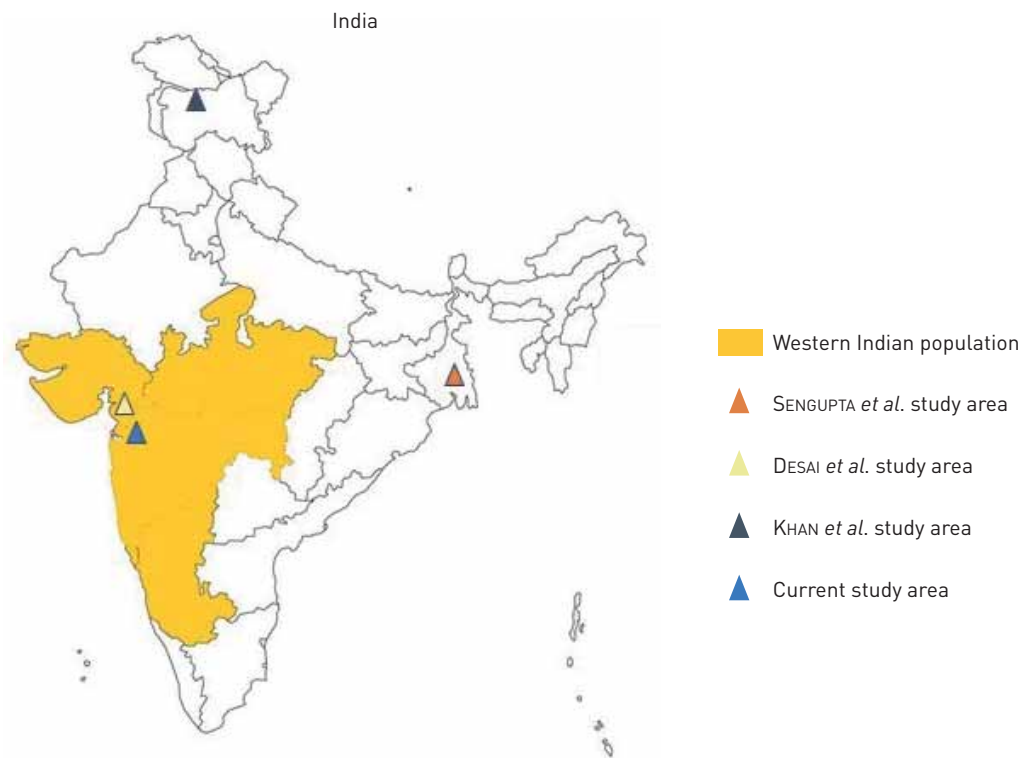


FIGURE 1 India map showing geographic distribution of the Western population in India and the location of previous studies conducted in India.

Methodology

We analysed spirometry data collected in two population surveys conducted in the Vadu Health and Demographic Surveillance System (Vadu HDSS) population of Vadu Rural health Programme, KEM Hospital Research Centre, Pune India in the period 2008 to 2012. One of the two original surveys aimed to estimate the prevalence of OADs in a rural adult population from Western India [4]; the other aimed to study the phenotypic characterisation of the adult healthy population using Ayurveda science (system of medicine with historical roots in the Indian subcontinent) [10]. Here we have used the data to develop reference equations for lung function parameters for this population.

Ethical consideration

Ethical approval for the secondary data analysis was obtained from the Ethics Committee of KEM Hospital Research Centre, Pune, who also provided approvals for the primary surveys in which the spirometry data were collected. All participants had provided written informed consent at the time of data collection for the original survey.

Study population

The study area was Vadu HDSS which is located in Pune district of Maharashtra State of Western India. Western India also includes Gujarat, Madhya Pradesh, Goa and Karnataka states. Although this distribution is based on geography (figure 1) and not ethnicity of the population the study population is distinct in terms of physique, nutrition, climate and culture compared with southern, northern and eastern populations [11–15]. There is some variation within the region in terms of local dialects and staple food as well as differences in terrain, rainfall and temperature. Defined geographically, the population of these states is approximately 20% of the total Indian population.

Study procedures

13 600 adults (men and women, aged 18 years and above) were selected for the two surveys using random number schemes from the Vadu HDSS sampling frame out of the total adult population of Vadu HDSS. We ensured that no participant was part of both population surveys. We had spirometry data from 2598 adults: 2083 able to provide spirometry from the OAD study and 515 randomly selected from the non-respiratory study. We performed spirometry using calibrated EasyWare NDD spirometers. Spirometry was performed by study staff from Vadu HDSS who were trained for conducting spirometry using the American Thoracic Society (ATS) and European Respiratory Society (ERS) guidelines [16]. Quality checks of the spirometry data generated were performed by S. Salvi; (respiratory physicians Chest Research Foundation, Pune, India). Height and weight of the study participants were assessed by trained field workers using calibrated stadiometer and adult weighing scale to a precision of 0.1 cm or 0.1 kg, respectively. Age and ethnicity were reported by the participants. Age was recorded in completed years as exact birth dates for some older participants were not available.

Creation of dataset for current analysis

By virtue of routine surveillance of the population, we had detailed medical, personal, biomass usage, smoking status and socio-demographic information on all sampled individuals. Based on these data, we classified participants as healthy or non-healthy. Healthy participants were defined as those who were never smokers, no occupational exposure to known hazards and non-alcohol drinkers. They also had to be free of any self-reported illness/disease conditions or respiratory symptoms in the last 12 months, no previous history of any lung disease and no previous hospitalisation for respiratory illness. Information to define healthy participants were collected using the Burden of Obstructive Lung Disease Initiative (BOLD) study questionnaire (refer to supplementary file 1 for BOLD questions used to define “healthy”) [17]. We also excluded anyone with a FEV₁/FVC ratio of <0.5 (even if they had no respiratory symptoms and were apparently disease free) as this represents clear “non-healthy” lung function by comparison with any reference guidelines.

Statistical analysis

We performed the analysis using the Stata v15 software and R software version 3.5.3 [18]. Descriptive statistics were performed for all demographic and spirometry parameters. Following QUANJER *et al.* [2] and COLE *et al.* [19], we applied the GAMLSS methodology to fit flexible statistical models to each of the outcomes FEV₁, FVC and FEV₁/FVC; with age and/or height included as predictors. No assumption was made regarding the similarity of models across sex strata, and separate models were fitted within male and female groups (*i.e.* we used a completely stratified analysis). Full details about the statistical models fitted can be found in supplementary file 2.

Models were fitted using the GAMLSS package [20] in R software version 3.5.3 [18]. The model predicted values, 5th centile values and 2.5th centile values, were each presented graphically and in look-up tables. Akaike's Information Criterion and the Schwarz Bayesian Criterion were calculated to assess model fit. Cox and Snell's pseudo R-squared values allowed us to assess the improvement in model goodness-of-fit compared with a null model with no covariates.

Internal validation of the derived models was accomplished using a bootstrap method. The regression coefficient (calibration slope) for the relationship between the bootstrap model predicted values and observed outcome values was calculated. We also calculated mean values of 95% limits of agreement for the bootstrap values compared to observed outcome values.

In addition, lower limit of normal values (2.5th and 5th centiles) were computed corresponding to the GLI Ethnicity Code 5 (other/mixed) models using supplementary material from QUANJER *et al.* [2]. We then calculated the proportion of our population sample that was below the GLI model-derived 2.5th and 5th centiles, and compared these proportions to those using our Western Indian population models. We also calculated the proportion of participants below the GLI centile limits who were then reclassified above and below our new centile limits.

Results

From 2598 individuals, we had unique, acceptable and reproducible spirometry data for 2007 participants. Spirometry data of 591 (22.8%) participants were excluded as they did not meet ATS/ERS criteria. After screening for healthy criteria, we had 1258 records available for final analysis (694 females and 564 males). The demographic characteristics of the study population and descriptive analysis of the key spirometry parameters (FEV₁, FVC and FEV₁/FVC) are presented in table 1.

Derived prediction equations

Prediction equations were developed using the GAMLSS package as recommended by the GLI initiative [2]. According to the Schwartz Bayesian Criterion, linear regression models were found to be sufficient for modelling the outcomes in males, although not for modelling FEV₁ and FVC in females. The model predicted values, 5th centile and 2.5th centile values (lower limit of normal) are shown graphically in a line plot for FEV₁/FVC (figure 2). We used contour plots for the models that included both age and height (FEV₁ and FVC) (supplementary file 3) because the results are in three dimensions. The model predicted values, 5th centile and 2.5th centile (lower limit of normal) values are shown in supplementary tables S1 to S18. The guideline-recommended fixed ratio of 0.7 [1] is shown on the FEV₁/FVC plots for reference (figure 2).

The prediction model for FEV₁/FVC in males and females were linear regression models with age as a single predictor. FEV₁/FVC appears to have a very strong age effect, as found in over 50 other studies [21]. This suggests that using a fixed 0.7 cut-off value regardless of age is not appropriate and will tend to result

TABLE 1 Demographic and spirometry characteristics of the study population (n=1258; male: n=564 [44.8%]; female: n=694 [55.2%])

Characteristics	Sex	Minimum	Maximum	Mean±sd
Age years	Male	19	88	40.9±13.3
	Female	19	79	38.9±11.5
Weight kg	Male	34	100	61.6±11.5
	Female	30	85	50.1±9.4
Height cm	Male	137	185	165.3±6.7
	Female	137	173	152.5±5.8
BMI kg·m ⁻²	Male	13.7	38.4	22.5±3.7
	Female	14.0	36.2	21.5±3.8
FEV ₁ L	Male	1.4	4.6	2.8±0.5
	Female	0.9	3.9	2.0±0.4
FVC L	Male	1.8	5.2	3.4±0.6
	Female	0.9	4.6	2.4±0.4
FEV ₁ /FVC	Male	0.53	0.97	0.81±0.069
	Female	0.56	1.00	0.83±0.068

BMI: body mass index; FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity.

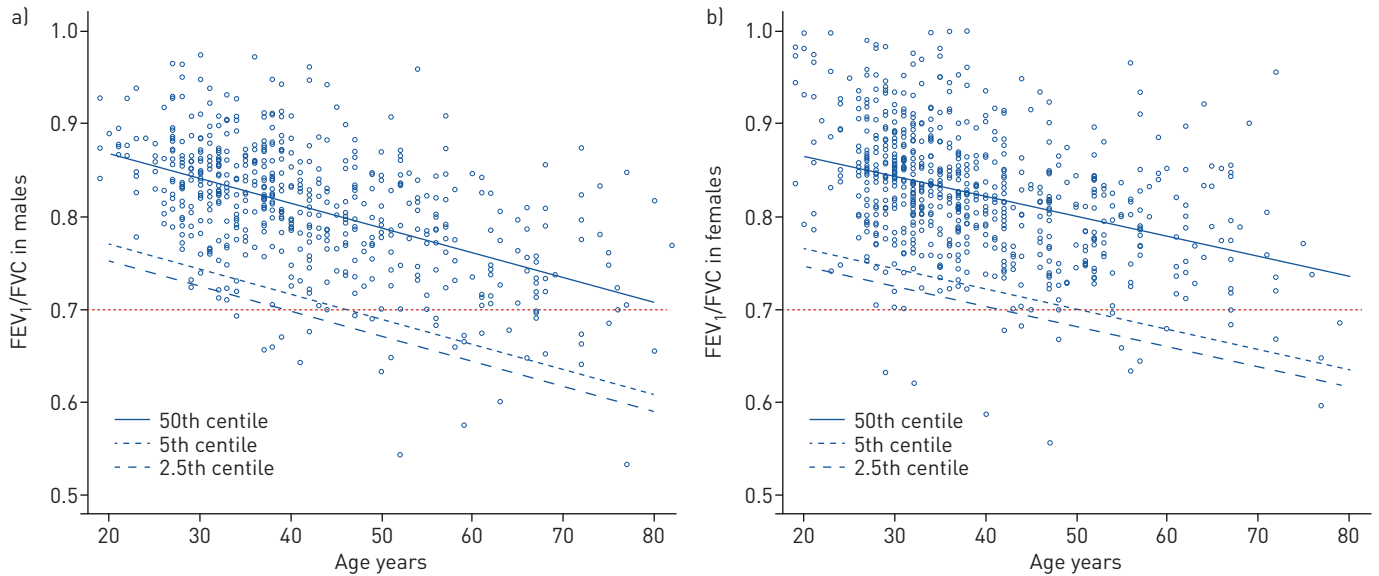


FIGURE 2 a) Line plot showing raw data points and 2.5th, 5th and 50th centile values for forced expiratory volume in 1 s (FEV_1)/forced vital capacity (FVC) for males with respect to age in years. b) Line plot showing raw data points and 2.5th, 5th and 50th centile values for FEV_1 /FVC for females with respect to age in years.

in under-diagnosis of people under the age of 40 years and over-diagnosis of people over 45 years (supplementary file 3). Prediction equations derived using the GAMLSS analysis are shown in table 2.

Overall, the bootstrap results confirmed the validity of the models (table 3). The mean calibration slopes are very close to 1 for all models except possibly for FVC in females and FEV_1 /FVC in females. The pseudo R-squared values for these models show a similar pattern; with relatively low R-squared values for FVC and FEV_1 /FVC in females (table 4). However, the alternative models we considered did not show any clear improvement in model performance (models indicated with + in tables 3 and 4).

Comparison with other studies

The comparison of predicted values (FEV_1 , FVC and FEV_1 /FVC) generated from this study with other previously published Indian studies and with GLI values (Ethnicity Code 1: Caucasian population; and

TABLE 2 Prediction equations for various spirometry parameters for subjects by using GAMLSS

Parameter	Sex strata	Model	Derived equation
FEV_1 L	Female	Inverse Gaussian with age and height	$\mu = \exp\{-0.6439556 - 0.0066921 \times \text{age} + 0.0104873 \times \text{height}\}$ $\sigma = \exp\{-2.911064 + 0.017174 \times \text{age}\}$
FVC L	Female	Gamma with age and height	$\mu = \exp\{-0.5590106 - 0.0040786 \times \text{age} + 0.0105529 \times \text{height}\}$ $\sigma = \exp\{-2.235059 + 0.007378 \times \text{age}\}$
FEV_1 /FVC	Female	Linear regression with age only	$\mu = 0.905877 - 0.002132 \times \text{age}$ $\sigma = 0.060512$
FEV_1 L	Male	Linear regression with age and height	$\mu = 0.402779 - 0.021695 \times \text{age} + 0.019853 \times \text{height}$ $\sigma = 0.413964$
FVC L	Male	Linear regression with age and height	$\mu = -0.038933 - 0.016466 \times \text{age} + 0.025120 \times \text{height}$ $\sigma = 0.473403$
FEV_1 /FVC	Male	Linear regression with age only	$\mu = 0.922257 - 0.002695 \times \text{age}$ $\sigma = 0.059375$

FEV_1 : forced expiratory volume is 1 s; FVC: forced vital capacity.

TABLE 3 Bootstrap validation results

Outcome	Model	Mean calibration slope [#]	Mean lower 95% LoA [¶]	Mean upper 95% LoA [¶]	Mean width of 95% LoA [¶]
FEV₁ in females	Inverse Gaussian with age and height	0.90	-0.65	0.60	1.25
FEV₁ in females⁺	Linear regression with age and height	0.92	-0.65	0.60	1.25
FEV₁ in males	Linear regression with age and height	0.99	-0.86	0.83	1.68
FVC in females	Gamma with age and height	0.88	-0.77	0.69	1.46
FVC in females⁺	Linear regression with age and height	0.88	-0.76	0.69	1.46
FVC in males	Linear regression with age and height	0.99	-0.99	0.94	1.93
FEV₁/FVC in females	Linear regression with age only	1.13	-12.22	12.79	25.01
FEV₁/FVC in females⁺	Linear regression with age and height	1.13	-12.22	12.79	25.01
FEV₁/FVC in males	Linear regression with age only	1.03	-11.61	11.82	23.43
FEV₁/FVC in males⁺	Linear regression with age and height	1.03	-11.61	11.82	23.43

[#]: The regression coefficient for the relationship between the bootstrap model predicted values and observed outcome values, averaged across 999 bootstrap datasets; [¶]: based on 95% limits of agreement (LoA) for the agreement between bootstrap model predicted values and observed outcome values, averaged across 999 bootstrap datasets; ⁺: standard alternative models (linear regressions with age and height) for comparison purposes.

Code 5: other/mixed population) for both males and females is illustrated in figure 3. Regarding the comparison of lower limit of normal values between this study and GLI Ethnicity Code 5 (other/mixed population), a much higher proportion of the study population sample were considered to be below the GLI limits compared to when using the limits generated from this study for FEV₁ and FVC (table 5). Indeed, a high proportion of the sample participants were reclassified to be above the limits based on our new models for FEV₁ and FVC when compared with the GLI limits, and a small proportion for FEV₁/FVC (table 5). This suggests that GLI Ethnicity Code 5 has the potential to over diagnose respiratory disease in the Western Indian population, particularly in terms of FEV₁ and FVC. This finding is confirmed by comparison of the differences between the absolute values of the model residuals (differences between observed and predicted values) which shows that the absolute model residuals of GLI Code 5 are between 0.25 and 0.27 greater on average for FEV₁ and FVC compared with the models used in this study (figures A and B in supplementary file 4). Bland-Altman plots also showed clear differences between the predicted values (figure 4a and b), particularly for those participants with relatively high values of FEV₁ and FVC.

Discussion

Clinicians rely on accurate reference values for lung function parameters to define normal lung function and to identify values that suggest disease; most commonly asthma or COPD. Normal ranges for lung

TABLE 4 Model diagnostics

Outcome	Model	n	AIC	SBC	Cox and Snell's pseudo R-squared [#]	Scale parameter σ^{\ddagger}
FEV₁ in females	Inverse Gaussian with age and height	585	257	279	0.352	0.075 to 0.211
FEV₁ in females⁺	Linear regression with age and height	585	292	310	0.301	0.309
FEV₁ in males	Linear regression with age and height	500	545	562	0.398	0.414
FVC in females	Gamma with age and height	585	431	453	0.257	0.123 to 0.192
FVC in females⁺	Linear regression with age and height	585	446	463	0.245	0.352
FVC in males	Linear regression with age and height	500	679	696	0.287	0.473
FEV₁/FVC in females	Linear regression with age only	585	-1616	-1602	0.139	0.061
FEV₁/FVC in females⁺	Linear regression with age and height	585	-1614	-1596	0.139	0.060
FEV₁/FVC in males	Linear regression with age only	500	-1399	-1386	0.256	0.059
FEV₁/FVC in males⁺	Linear regression with age and height	500	-1397	-1380	0.256	0.059

AIC: Akaike's Information criterion; SBC: Schwarz Bayesian criterion. Lower AIC or SBC indicates better model performance. AIC and SBC cannot be used to compare across models with different outcomes or population subgroups. [#]: Values are equivalent to the standard model R-squared if model is linear regression; [¶]: for some models, the scale parameter is given as a range because it varies with age; can be interpreted as the residual standard error for linear regression models; ⁺: standard alternative models (linear regressions with age and height) for comparison purposes; these generally have poorer goodness-of-fit.

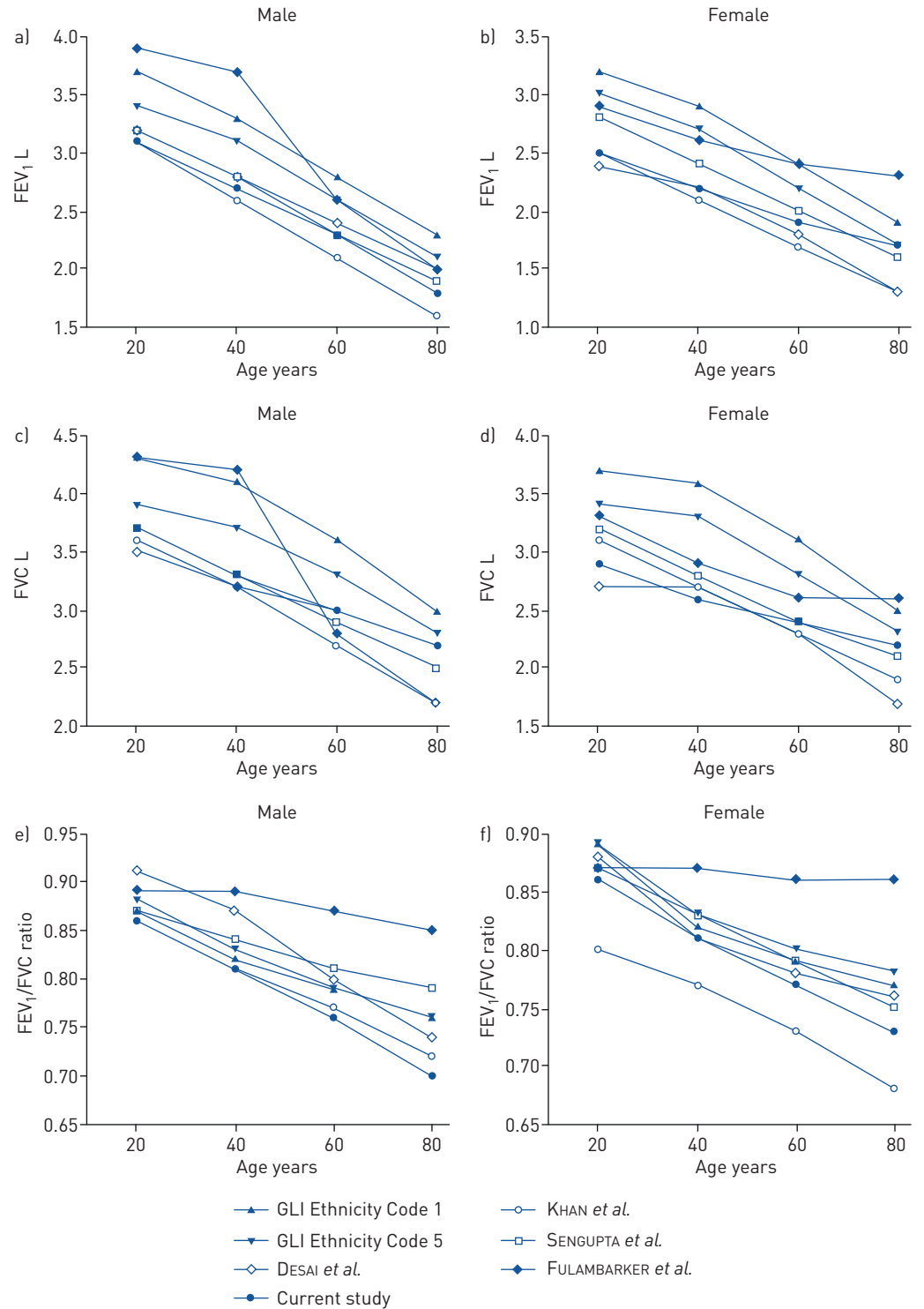


FIGURE 3 Comparison of spirometry predictive values obtained from this study with other published studies from India and Global Lung Function Initiative (GLI) [Ethnicity Code 1: White Caucasian, Code 5: other/mixed] at age 20, 40, 60 and 80 years for height 160 cm in both males and females.

function are dependent on a number of factors such as age, sex, height and also ethnicity. Our findings, from population surveys of healthy individuals, presents prediction equations and predictive values for the Western Indian adult population. Using GAMLSS models, it was found that age and height were the main predictors of the FEV₁ and FVC spirometry parameters in both sexes. For FEV₁/FVC, height was not a significant predictor of outcome and so was not included in our models.

TABLE 5 Comparison of our models with Global Lung Function Initiative (GLI) Code 5 (other/mixed) models according to proportion of our population sample that is below the model-derived 2.5th and 5th centiles, and proportion of participants reclassified after using our models

Outcome	Centile	Sex	Proportion of sample below GLI limit (GLI code 5 other/mixed equations)	Proportion of sample below new limit (our Western Indian models)	Proportion reclassified above new limit [#]
FEV₁	5th	Male	33% (185/564)	7% (38/564)	26% (147/564)
		Female	45% (312/694)	7% (53/694)	37% (259/694)
	2.5th	Male	21% (121/564)	3% (15/564)	19% (106/564)
		Female	31% (213/694)	5% (33/694)	26% (180/694)
FVC	5th	Male	34% (191/564)	5% (29/564)	29% (162/564)
		Female	41% (287/694)	7% (49/694)	34% (238/694)
	2.5th	Male	24% (135/564)	3% (18/564)	21% (117/564)
		Female	30% (207/694)	5% (32/694)	25% (175/694)
FEV₁/FVC	5th	Male	8% (44/564)	6% (33/564)	2% (11/564)
		Female	7% (49/694)	4% (30/694)	3% (19/694)
	2.5th	Male	4% (21/564)	3% (16/564)	1% (5/564)
		Female	3% (24/694)	2% (15/694)	1% (9/694)

[#]: Note that no participants (0%) were reclassified below the new limit across all models. FEV₁: Forced expiratory volume in 1 s; FVC: forced vital capacity.

Comparison with data from other surveys

When compared with the data from other Indian studies and the GLI initiatives, the Western Indian adult population had lower lung volumes on average compared with the European-American population [2] or to populations from Northern India [9] (Refer to figure 3 for comparison) [2, 22, 23]. Indeed, we found that many of our participants would have been classified as being below the lower limit of normal using the GLI Ethnicity Code 5 (other/mixed populations) limits. This is a significant finding as it indicates that using reference predictive values derived from non-Indian (or Northern Indian) populations will over diagnose some lung conditions in the Western Indian adult population, though the FEV₁/FVC ratio which identifies OADs is similar to GLI. Figure 1 indicates the location of previous studies in India.

In contrast to other studies, we did not find that height was a significant predictor of FEV₁/FVC. This may be because we did not have data from children, restricting the range of the explanatory variable (height). We also observe that the relationships between height and outcome were very similar for FEV₁ and FVC individually (table 2), and so when we take the ratio there is no residual effect of height.

The question arises whether the smaller lung volumes observed in our study compared to other populations, are due to inherent ethnic differences or are the result of environmental influences on the lungs. For our analysis, we defined the healthy population as rigorously as possible by selecting individuals with no respiratory symptoms, no apparent illnesses and no exposure to smoking or known occupational hazards. This excluded more elderly women than elderly men (see demography in table 1) possibly because the ubiquitous exposure to biomass fuel preferentially affects women (and girls from a young age) as they are responsible for cooking. We also excluded individuals with clearly abnormally low FEV₁/FVC ratios even if they had no symptoms or history of disease. However, clinical examinations were not performed to confirm reported health status. Our findings are derived from adults, and from one rural region of India limiting generalisability.

There are multiple factors that influence lung development (e.g. exposure to biomass fuel, vehicular pollution, nutritional deficiencies, (passive) smoking, sedentary lifestyle, obesity) which will play out differently in rural and urban communities. Children living in rural communities have smaller lung volumes than their urban counterparts [21] but the situation in adults may be less clear cut as young people from rural communities move into cities and experience a lifetime of exposure to environmental factors. However, comparison with a small study in a slightly different population (from Mumbai) showed similar lung volumes to our rural study [5]. In addition, a survey from the USA confirmed similarly reduced lung volumes in Asian Indians (specific ethnic groups were not identified) living in Chicago, IL [24], and assumed not to have been exposed to biomass fuel or high levels of pollution (figure 3 illustrates the comparison). Our findings may thus have relevance to physicians caring for migrant Western Indian populations in USA, Europe (especially UK), and the Gulf States, albeit with the caveat that the US study did not report the ethnic group of the Asian Indian migrants or duration of their residence in the USA.

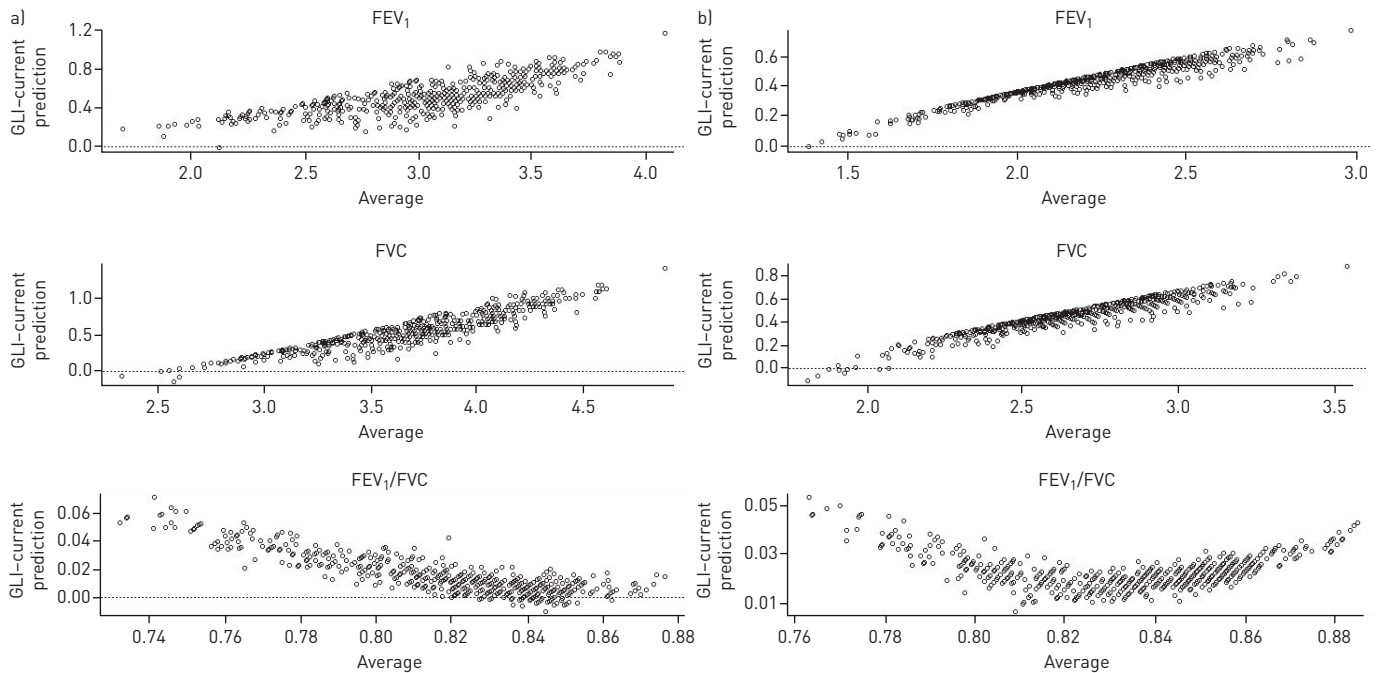


FIGURE 4 a) Bland-Altman plots showing the comparison of spirometry predictive values obtained from this study with Global Lung Function Initiative (GLI) Code 5 (other/mixed): males. b) Bland-Altman plots showing the comparison of spirometry predictive values obtained from this study with GLI Code 5 (other/mixed): females.

Our study has many strengths. Our sample size of 1258 healthy participants is four-times that of previous studies in the Western Indian population, though it remains small for creating generalisable equations. This is also the first study in the Indian population which has used GAMLSS models to derive reference equations and predictive values. This meant that we were not restricted to certain types of models (*e.g.* linear regression models) with their implicit assumptions. Instead, our aim was to take a flexible parametric approach *via* GAMLSS models in order to maximise the fit of the models to the data and thereby improve predictions. Nevertheless, for the lung function outcomes in males, our final models were, in fact, linear regression models with age and/or height, confirming the validity of using such models, at least in the Western Indian male population. We also performed internal validation of the statistical models by calculating model diagnostics and using bootstrap methods. Pseudo R-squared values were low, particularly for FEV₁/FVC in females. Note that this does not suggest that the models were inferior: our thorough modelling approach means that it is unlikely that better fitting models are possible, at least based on the data we collected. It does mean, however, that there is still a large amount of unexplained variability in response and using the models to try to make exact predictions is inadvisable (especially for FEV₁/FVC). This highlights the importance of using both the lower limit of normal values as well as predictive values when assessing lung function.

We realise that to define those who were truly healthy, we would need to examine the individuals in greater detail. Even then, it is difficult to confirm if someone is truly healthy with the complete absence of any pathology. We thus made best use of the available data to define healthy individuals.

Conclusion

Reflecting the ethnically diverse populations in India, Western Indian individuals from rural populations have small lung volumes relative to European-Americans or populations from Northern Indian populations. Use of GLI normal values are thus potentially misleading in this population and may result in overdiagnosis of lung disease. We recommend that the values and equations generated from this study should be used by physicians in their routine practice for diagnosing the disease condition and its severity in adults from Western Indian populations. There are large Western Indian migrant communities in Europe (principally in the UK) and the USA for whom these data may also be relevant. We will be sharing our data with the GLI network.

Acknowledgments: We gratefully acknowledge all the study participants for their participation in the study. We also acknowledge Field Research Assistants of Vadu HDSS for data collection and performing spirometry of study participants. RP is partly supported in this work by NHS Lothian *via* the Edinburgh Clinical Trials Unit.

Author contributions: S. Juvekar and H. Pinnock led the development of the study. D. Agarwal wrote the first draft of the manuscript which was critically reviewed and refined by R. Parker, S. Juvekar, H. Pinnock, S. Roy, D. Ghorpade, S. Salvi, and P. Khatavkar. R. Parker performed the statistical analysis. RESPIRE UMC members provided advice and contributed to discussions. All authors approved the final version.

The RESPIRE collaboration: The University of Edinburgh, Edinburgh, UK: Harry Campbell, Harish Nair, Steve Cunningham, Monica Fletcher, Liz Grant, Aisha Holloway, Aziz Sheikh, Pam Smith; The Allergy & Asthma Institute, Islamabad, Pakistan: Osman Yusuf, Shahida Yusuf; Maternal Neonatal and Child Health Research Network, Islamabad, Pakistan: Hana Mahmood; University of Malaya, Malaysia: Wong LP; KEM Hospital Research Centre, Pune, India: Anand Kawade; Aga Khan University, Karachi, Pakistan: Sajid Bashir Soofi; Christian Medical College, Vellore, India: Rita Isaac; Victoria University of Wellington, New Zealand: Colin Simpson.

Conflict of interest: R. Parker has nothing to disclose. H. Pinnock reports grants from NIHR Global Health Research (ref: 16/136/109, NIHR Global Respiratory Health (RESPIRE) Unit), during the conduct of the study. S. Roy reports grants from NIHR Global Health Research (ref: 16/136/109, NIHR Global Respiratory Health (RESPIRE) Unit), during the conduct of the study. D. Ghorpade has nothing to disclose. S. Salvi has nothing to disclose. P. Khatavkar has nothing to disclose. S. Juvekar reports grants from NIHR Global Health Research (ref: 16/136/109, NIHR Global Respiratory Health (RESPIRE) Unit), during the conduct of the study. D. Agarwal reports grants from NIHR Global Health Research (ref: 16/136/109, NIHR Global Respiratory Health (RESPIRE) Unit), during the conduct of the study.

Support statement: This study funded by the NIHR Global Health Research Unit in Respiratory Health (RESPIRE) at the Usher Institute of Population Health Sciences and Informatics. RESPIRE was commissioned by the National Institute of Health Research using Official Development Assistance (ODA) funding. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. Funding information for this article has been deposited with the Crossref Funder Registry.

References

- 1 Global Initiative for Chronic Obstructive Lung Disease. GOLD Report 2019. 2019: 1–155. <https://goldcopd.org/> Date last accessed: June 29, 2019.
- 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 Mukherjee S, Banerjee G, Das D, *et al.* Occurrence of COPD in patients with respiratory allergy: a clinico-spirometric evaluation in a tertiary hospital Kolkata. *J Clin Diagnostic Res* 2017; 11: 11–13.
- 4 Amaral AFS. Airflow obstruction and use of solid fuels for cooking or heating: BOLD results. *AJRCCM* 2012; 44: 1–43.
- 5 Desai U, Joshi JM, Chhabra SK, *et al.* Prediction equations for spirometry in adults in western India. *Indian J Tuberc* 2016; 63: 176–182.
- 6 Dasgupta A, Ghoshal AG, Mukhopadhyay A, *et al.* Reference equation for spirometry interpretation for Eastern India. *Lung India* 2015; 32: 34–39.
- 7 Nayak PK, Satpathy S, Manjareeka M, *et al.* Normal spirometric standards in young adult Indian population. *J Basic Clin Physiol Pharmacol* 2015; 26: 321–325.
- 8 Chhabra SK, Kumar R, Mittal V. Prediction equations for spirometry for children from northern India. *Indian Pediatr* 2016; 53: 781–785.
- 9 Saleem S, Shah S, Gailson L, *et al.* Normative spirometric values in adult Kashmiri population. *Indian J Chest Dis Allied Sci* 2012; 54: 227–233.
- 10 Chauhan NS, Pandey R, Mondal AK, *et al.* Western Indian rural gut microbial diversity in extreme Prakriti endo-phenotypes reveals signature microbes. *Front Microbiol* 2018; 9: 1–12.
- 11 Karve IK. Maharashtra, Land and Its People. Mumbai, Directorate of Government Printing, Stationery and Publications, 1968.
- 12 Basu A, Mukherjee N, Roy S, *et al.* Ethnic India: A genomic view, with special reference to peopling and structure. *Genome Res* 2003; 13: 2277–2290.
- 13 Majumder PP. People of India: biological diversity and affinities. *Evol Anthropol* 1998; 6: 100–110.
- 14 Karve IK. Hindu Society: An Interpretation. Pune, Deshmukh Prakashan, 1968.
- 15 Beteille A. The Indian heritage—A sociological perspective. In: D Balasubramian, N Appaji Rao, eds. The Indian human heritage. Hyderabad, University Press, 1998.
- 16 Pellegrino R, Viegi G, Brusasco V, *et al.* Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.
- 17 Buist AS, Vollmer WM, Sullivan SD, *et al.* The Burden of Obstructive Lung Disease Initiative (BOLD): rationale and design. *J Chronic Obstr Pulm Dis* 2005; 2: 277–283. www.ncbi.nlm.nih.gov/pubmed/17136954.
- 18 R Core Team. R: A language and environment for statistical computing. 2019 <https://www.r-project.org/>.
- 19 Cole TJ, Stanojevic S, Stocks J, *et al.* Age- and size-related reference ranges: a case study of spirometry through childhood and adulthood. *Stat Med* 2009; 28: 1999–2011.
- 20 Stasinopoulos DM, Rigby RA. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *J Stat Softw* 2007; 23: 1–46.
- 21 Swanney MP, Ruppel G, Enright PL, *et al.* Using the lower limit of normal for the FEV1/FVC ratio reduces the misclassification of airway obstruction. *Thorax* 2008; 63: 1046–1051.
- 22 Knudson RJ, Burrows Lebowitz B, Holberg CJ. Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Respir Dis* 1983; 127: 725–734.
- 23 Crapo RO, Morris AT, Gardner RM. Spirometric values using techniques and equipment that meet ATS recommendations. *Am Rev Respir Dis* 1981; 123: 659–664.
- 24 Fulambarker A, Copur AS, Javeri A, *et al.* Reference values for pulmonary function in Asian Indians living in the United States. *Chest* 2004; 126: 1225–1233.