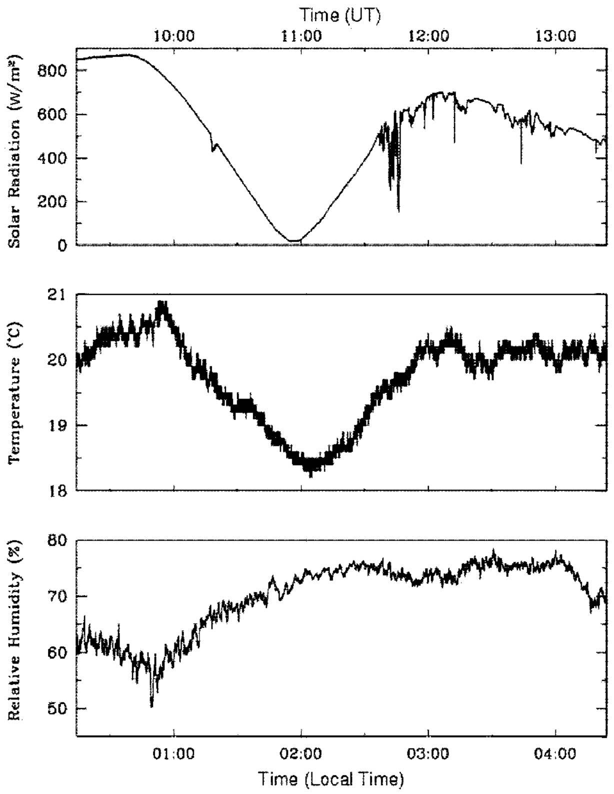

Figure 5. Plot showing variation of solar radiation, temperature and relative humidity (top to bottom) in the earth during the eclipse, covering the whole event duration from its first to fourth contacts.

ring 10–30 min after totality. However, Fernadez *et al.*[6] in their 1994 total eclipse observations found a surface temperature decrease of 3°C, with the lowest value occurring about 7 min after totality. The present observations also confirm similar changes in the atmospheric parameters. The time lag between the temperature minimum and the time of totality may be interpreted in terms of the thermal inertia of air and ground[6]. Due to its dependence on temperature, relative humidity increases as a consequence of decrease in temperature (third panel, Figure 5).

1. Dwivedi, B. N., The solar corona. In *Lectures on Solar Physics* (eds Antia, H. M., Bhatnagar, A. and Ulmschneider, P.), Lecture Notes in Physics, Springer-Verlag, 2003, vol. 619, pp. 281–298.
2. Koutchmy, S., Zhugzhda, Ia. D. and Locans, V., *Astron. Astrophys.*, 1983, **120**, 185–191.
3. Singh, J. *et al.*, Detection of short-period coronal oscillations during the total solar eclipse of 24 October 1995. *Sol. Phys.*, 1997, **170**, 235–252.
4. Cowsik, R. *et al.*, Short-period intensity oscillations in the solar corona observed during the total solar eclipse of 26 February 1998. *Sol. Phys.*, 1999, **188**, 89–98.
5. Fernandez, W., Castro, V. and Hidalgo, H., Air temperature and wind changes in Costa Rica during the total solar eclipse of July 11, 1991. *Earth, Moon, Planets*, 1993, **63**, 133–147.
6. Fernandez, W. *et al.*, Changes in meteorological variables in Coronel Oviedo, Paraguay, during the total solar eclipse of 3 November 1994. *Earth, Moon, Planets*, 1996, **74**, 49–59.
7. Jain, S. L. *et al.*, Measurements of various atmospheric parameters during the solar eclipse. *Terr. Atmos. Ocean. Sci.*, 1997, **8**, 371–384.

ACKNOWLEDGEMENTS. We thank the anonymous referee for constructive comments which improved the scientific contents of the paper. We thank the staff of Solar, Mechanical, Electronics and Optics sections of ARIES, Nainital for help during various stages of this expedition. We also thank Dr Atila Özgüç and his team from Kandilli Observatory, Bogazici University, Istanbul, Turkey for helping to carry out observations of the solar eclipse at Manavgat, Turkey. We thank Prof. Jagdev Singh for fruitful discussions.

Received 15 March 2007; revised accepted 13 August 2007

Prediction models for peak expiratory flow rates in North Indian male population based on ordinary and weighted least square estimation

Neeraj Mathur[1], Subhodh Kumar Rastogi[1,*], C. Kesavachandran[1] and G. G. Agarwal[2]

[1]Epidemiology Division, Industrial Toxicology Research Centre, MG Road, Lucknow 226 001, India
[2]Department of Statistics, Lucknow University, Lucknow 226 001, India

The present study was carried out to establish physiological norms (best regression equation) to predict peak expiratory flow values among the North Indian male population with a statistically appropriate model. The study aims to establish the best statistically sound multiple regression model for predicting peak expiratory flow rate in the North Indian healthy population considering age, height and weight as predictor variables. One hundred and thirty-seven normal male subjects aged between 20 and 69 years, who had come from different parts of Lucknow to attend a science exhibition, India were selected for the study. The ordinary least square multiple regression model was used for the study. Residuals in this model were heteroskedastic. Therefore, the model proposed by Prasad *et al.* and the weighted least square models were also used for the study. All the models were compared statistically. The proposed weighted least square model was found to be best fitting model, which had minimum residual standard deviation, maximum explained variation and most precise regression coefficient estimates.

Keywords: Least square model, male population, peak flow rate, prediction models.

PEAK flow rate (PEFR) is one of the useful and simple parameters for assessing the lung function status in gen-

*For correspondence. (e-mail: subhodhrastogi@yahoo.com)

eral population and also for making a diagnosis and monitoring treatment of patients with bronchial asthma and chronic obstructive lung disease. Many studies on PEFR in normal men and women belonging to different populations representing various racial and ethnic groups have been carried out in India[1–9] and abroad[10–15]. Multiple regression models considering PEFR as a dependent variable, and age, height and weight as independent variables were fitted and globally compared by Mathur *et al.*[16], representing various racial and ethnic groups. In almost all the studies, the regression coefficients of age, height and weight were calculated using ordinary least square (OLS)[5–14] approach, without examining the homoskedasticity of residuals with respect to predictor variables, which is an integral part of model-building[17–19]. The residuals are generally heteroskedastic, i.e. the standard deviation of the error term is not generally constant over all values of the predictor variables. The assumption of constant standard deviation, however, clearly does not hold in every modelling application. There is only one study done by Prasad *et al.*[20], where the heteroskedasticity was examined. They observed empirically that the error variance was not constant, but tends to increase with age.

The present study was carried out to establish physiological norms (best regression equation) to predict peak expiratory flow values among the male population from the northern part of India using a statistically appropriate model. The OLS model, the model proposed by Prasad *et al.*[20] and weighted least square models were compared by examining the residuals, comparing the standard error of the regression coefficients of age, height and weight, coefficient of determination and residual standard deviations of each model.

The study was carried out on 137 male, non-smoker subjects aged between 20 and 68 years. The subjects were selected from different parts of Lucknow, India, who had attended science exhibition organized by the CSIR. Data were collected by the authors (S.K.R. and C.K.), who specialize in clinical and pulmonary physiology.

A detailed history and clinical examination of individuals was carried out. Subjects who were asymptomatic were chosen for the study. PEFR, defined as the maximum rate of expiratory air flow after a single deep inspiration, was measured using Wright's Peak Flow Meter (Clement Clerke, UK). Prior to measuring PEFR, clear instructions were given to each subject. The test was performed three times on each subject and the highest value among the three was selected for data computation. Age in completed years, standing height in nearest centimetre without shoes and body weight in kilograms with minimum clothing were recorded for each subject prior to PEFR recording.

OLS model (model 1): The multiple linear regression model considering PEFR as a dependent variable and age, height and weight as independent variables was fitted[18]. Homoskedastic of error variance (residuals) with respect to predictor (independent) variables was tested using graphical and statistical methods. In the graphical method, scatter diagrams of residuals and predictor variables were plotted. White's chi square[21] was calculated for testing the homoskedasticity of residuals.

Weighted least square models: The most important assumption in regression modelling was that each data point provides equally precise information about the deterministic part of the total process variation, i.e. standard deviation of the term is constant over all values of the predictor variable and this assumption does not hold in the OLS model. There are two basic approaches to obtain precise estimates of the regression coefficients.

(i) Transforming the data (model 2): The model of Prasad *et al.*[20] was used, where both independent and dependent variables were transformed by dividing the data of all variables by age, as it was observed empirically that the error variance tend to increase with age. The regression model was fitted considering PEFR/age as dependent variable and 1/age, height/age and weight/age as independent variables.

(ii) Using weights: In weighted least approach, the values of the constant and the regression coefficients were estimated which minimizes residual sum square

$$\mathrm{RSS} = \sum Wi(Y - \hat{Y})^2,$$

where Y is the observed PEFR, $\hat{Y}$ the predicted PEFR using OLS method, and Wi are the weights to be estimated. Two techniques for estimating the weights used were Residualized method and log likelihood method[22].

Residualized method (model 3): The residuals from OLS were squared, denoted by residsq. Residsq was regressed on independent variables (age, height and weight) to get the predicted values denoted by pre_2. All independent and dependent variables were transformed as follows

$$\mathrm{Wt}Xi = Xi/\sqrt{\mathrm{pre\_2}},$$

$$\mathrm{Wt}Y = Y/\sqrt{\mathrm{pre\_2}},$$

$$\mathrm{Wtsqroot} = 1/\sqrt{\mathrm{pre\_2}}.$$

Regression analysis through the origin was carried out considering the variable WtY as dependent, and Wtsqroot and WtXi as independent variables to get the desired regression model.

Log likelihood method (model 4): The weights were iteratively determined with the log likelihood procedure using the SPSS 10.0 software package considering age, height and weight as independent variables and observed PEFR as the dependent variable. Each variable (both dependent and independent) was considered separately as a source of heteroskedasticity, as a weight variable considering power values from –2 to +2 with an increment of 0.5. The log likelihood estimates for different models for each power value were calculated. The variable and power which gave maximum value of the likelihood ratio were chosen.

The statistical results were compared by computing the standard error of the constants and regression coefficients, residual standard deviation (RSD) and the explained variation (R^2) of each model. The model which had minimum standard error of regression was also considered as the criterion for the best regression model for predicting PEFR.

Table 1 shows age, height, weight and observed PEFR values (mean ± SD), Table 2 shows mean PEFR values in different age groups. The OLS model fitted was

$$\text{PEFR} = \beta_0 + \beta_1(\text{age}) + \beta_2(\text{height}) + \beta_3(\text{weight}) + \varepsilon. \qquad \text{(model 1)}$$

In this equation β_0, β_1, β_2 and β_3 are regression coefficients and ε is the random disturbance. The regression coefficient estimates, R^2 and RSD of this model are presented in Table 3. As indicated earlier, model-fitting is incomplete without regression diagnostics[23–25]. Residuals from the OLS method were found to be heteroskedastic, as scatter diagrams of residuals with predictor variables increased with increase in the predictor variables. White's chi square statistics for homoskedasticity was also found to be statistically significant ($\chi^2 = 13.0$, $P < 0.001$).

The model by Prasad *et al.*[20] was applied to the data. The model considered was as follows

$$(\text{PEFR/age}) = \beta_0(1/\text{age}) + \beta_1 + \beta_2(\text{height/age}) + \beta_3(\text{weight/age}) + \varepsilon/\text{age}. \qquad \text{(model 2)}$$

The regression analysis was performed with transformed variables. PEFR/age was considered as dependent variable, and 1/age, height/age and weight/age as independent variables. The equations of the fitted lines in terms of original variable were obtained after multiplying both sides of model 2 by age. The coefficients and the corresponding statistics are presented in Table 3.

The residualized regression model was fitted as follows

$$(\text{PEFR}/\sqrt{\text{pre\_2}}) = \beta_0(1/\sqrt{\text{pre\_2}}) + \beta_1(\text{age}/\sqrt{\text{pre\_2}}) + \beta_2(\text{height}/\sqrt{\text{pre\_2}}) + \beta_3(\text{weight}/\sqrt{\text{pre\_2}}), \qquad \text{(model 3)}$$

where pre_2 is as defined earlier. The coefficients and the corresponding statistics are presented in Table 3.

The log likelihood function (model 4) was found to be maximum with power value of 0.5, when the dependent variable PEFR was considered as the source of heteroskedasticity. So weight was estimated as

$$\text{Wt} = 1/(\text{observed PEFR})\text{^}0.5,$$

and the residuals and predicted values for each subject were calculated as

$$\text{Resd\_2} = (\text{Resd\_1}) * \sqrt{(\text{Wt})},$$

$$\text{Pred\_2} = (\text{Pred\_1}) * \sqrt{(\text{Wt})},$$

where Resd_1 and Pred_1 are the residuals and predicted values from OLS. The corresponding regression coefficients and other statistics are presented in Table 3. Model 3 was found to be the best-fitting model, which has maximum R^2 (0.907) and minimum RSD (0.008) and the regression constant and coefficients of age, height and weight were most precise (minimum standard error) compared to the other three models.

PEFR norms in the form of multiple linear regression model in male adults have been developed by various authors in India[1–5] and abroad[10–15], where regression constant and coefficients for age, height and weight were estimated using OLS method, without examining the residuals. A comparison of these models[16] indicated that explained variation (R^2) in PEFR by the explanatory variables ranged between 17 and 29% and the RSD between 39.9 and 91. In the present study using the OLS model (model 1), the values were almost in the same range ($R^2 = 35.3$, RSD = 67.4), though the sample size in most of the studies[1–5] was much more and almost the same in only two studies[12,14].

In most of the previous studies the emphasis was on finding the relationship between age, height, weight and PEFR. It was observed that the mean PEFR values declined with the advancement of age. Our study also demonstrated similar findings (Table 2). The present study takes residuals into consideration for model-building and compares the OLS model with the weighted least square models developed in the present study. As a consequence of this residual analysis, variation in PEFR explained by the fitted models improved to 69, 90.7 and 37.8% for models 2–4 respectively, compared to 35.3 for model 1.

Table 1. Mean values for age, height weight and PEFR ($n = 137$)

Parameter	Mean ± SD
Age (yrs)	46.0 ± 10.4
Height (cm)	161.1 ± 8.6
Weight (kg)	54.9 ± 13.0
PEFR (l/min)	292.6 ± 119.4

Table 2. PEFR values (mean ± SD) in different age groups

Age group (yrs)	*N*	PEFR ± SD
20–29	14	346.7 ± 104.9
30–39	19	329.3 ± 107.4
40–49	46	328.7 ± 116.2
50–59	60	248.6 ± 115.7
60+	7	197.0 ± 67.0

Table 3. Comparison of statistical results of different models

Model	Constant $\beta_0 \pm$ SE	Age $\beta_1 \pm$ SE	Height $\beta_2 \pm$ SE	Weight $\beta_3 \pm$ SE	R^2	RSD
Model 1	–253.118 ± 177.423	–4.750 ± 0.779	3.886 ± 1.212	2.400 ± 0.802	0.353	67.4
Model 2	–166.638 ± 165.441	–4.667 ± 0.713	3.177 ± 1.130	2.812 ± 0.764	0.690	0.037
Model 3 (proposed model)	–361.852 ± 153.558	–4.477 ± 0.708	4.644 ± 1.089	1.892 ± 0.761	0.907	0.008
Model 4	–250.011 ± 171.494	–4.980 ± 0.780	3.936 ± 1.177	2.063 ± 0.789	0.378	4.067

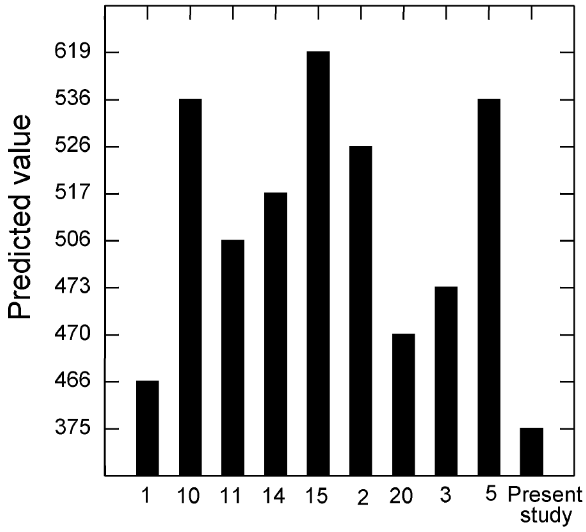

Figure 1. Predicted PEFR in different studies.

The linear additive model used in the present study is generally preferred over other models, because it is simple and convenient and the sum of squares explained by the model are adequately represented. Khosla[26], however, used a multivariate model to obtain indices of ventilatory functions in an industrial population from the steel industry. Cole[27] added the interaction term of age/height to the regression models relating Forced Vital Capacity (FVC) and Forced Expiratory Volume in 1 second (FEV_1) to physical characteristics. In a subsequent study, he preferred a proportional model. Malik *et al.*[1] used a polynomial model with age for relating PEFR to physical characteristics. Gupta and Mathur[3] reported that all these models had similar predictive capacity for pulmonary functions in a normal healthy population.

The regional and ethnic differences were also compared by plotting the predicted values of PEFR (Figure 1) from different studies carried out in India and abroad. Studies conducted in North India[1,3,20] showed almost similar values; however, a study by Rastogi *et al.*[5] in asymptomatic industrial workers from North India had higher values due to higher physical activity involved in work and use of safety gadgets to reduce the exposure of toxicants, resulting in better lung-function values. In the present study lowest PEFR values were observed compared to other studies carried out in India[1–3,5,20]. This may be due to decline of PEFR with age, and mean age, of the studied population was found to be higher compared to the other studies. The PEFR values of men of European descent[10,11,13–15] were greater than those of Indians. The differences were quite significant, which may be due to climatic conditions[28], and nutritional and genetic factors[29] leading to increased lung and chest cage size.

Therefore, the present study may be utilized by clinicians and pulmonary physiology researchers for the research and clinical diagnosis as prediction models for PEFR values in North Indian male population.

1. Malik, S. K., Jindal, V. and Bansal, S., Peak expiratory flow rates in healthy adults. *Indian J. Chest Dis.*, 1975, **17**, 166–171.
2. Kamat, S. R., Sarma, B. S. and Raju, V. R. K., Indian norms for pulmonary function. *J. Assoc. Physicians India*, 1977, **25**, 531–540.
3. Gupta, C. K. and Mathur, N., Statistical models relating peak expiratory rate of age, height, weight and sex. *J. Epidemiol. Commun. Health*, 1982, **36**, 64–67.
4. Mathur, N., Rastogi, S. K., Husain, T. and Gupta, B. N., Lung function norms in healthy working women. *Indian J. Physiol. Pharmacol.*, 1998, **42**, 245–251.
5. Rastogi, S. K., Mathur, N. and Clerk, S. H., Ventilatory norms in healthy industrial male workers. *Indian J. Chest Dis. Allied. Sci.*, 1983, **25**, 186–195.
6. Chowgule, R. V., Shety, V. M., Parmar, J. R., Lung function tests in normal Indian children. *Indian Pediatr.*, 1995, **32**, 185–191.
7. Vijayan, V. K., Reetha, A. M., Kuppurao, K. V., Venkatesan, P. and Thilakavathy, S., Pulmonary function in normal south Indian children aged 7–19 years. *Indian J. Chest Dis. Allied Sci.*, 2000, **42**, 147–156.
8. Raju, P. S., Prasad, K. V., Ramana, Y. V., Ahmed, S. K. and Murthy, K. J., Study on lung function tests and prediction equations in Indian male children. *Indian Pediatr.*, 2003, **40**, 705–711.
9. Raju, P. S., Prasad, K. V., Ramana, Y. V. and Murthy, K. J., Pulmonary function tests in Indian girls: Prediction equation. *Indian. J. Pediatr.*, 2004, **71**, 893–897.
10. Selby, T. and Read, J., Maximum expiratory flow rate in Australian adults. *Aust. Ann. Med.*, 1961, **10**, 49–53.
11. Ferris, B. G., Anderson, D. O. and Zickmantal, R., Prediction values for screening tests of pulmonary function. *Am. Rev. Respir. Dis.*, 1965, **91**, 252–261.
12. Johannson, Z. M. and Eramos, L. D., Clinical spirometry in normal Bantu. *Am. Rev. Respir. Dis.*, 1968, **97**, 585–589.
13. Elbute, E. A. and Femi Pearse, D., Peak flow rates in Nigerians. *Thorax*, 1971, **26**, 597–601.
14. Woolcock, A. J., Coleman, M. H. and Blackburn, C. R. B., Factors affecting normal values for ventilatory function. *Am. Rev. Respir. Dis.*, 1972, **106**, 692–709.
15. Gregg, I. and Nunn, J., Peak expiratory flow rates in normal subjects. *Br. Med. J.*, 1973, **3**, 282–284.

16. Mathur, N., Rastogi, S. K., Gupta, B. N. and Hussain, T., A global comparison of predicting equations on spirometry in male population. *Int. J. Epidemiol. Commun. Health*, 1990, **19**, 331–338.
17. Anscombe, F. J. and Tukey, J. W., The examination and analysis of residuals. *Technometrics*, 1963, **5**, 141–160.
18. Draper, N. R. and Smith, N., *Applied Regression Analysis*, John Wiley, New York, 1988.
19. Cook, R. D. and Weisberg, S., *Residuals and Influence in Regression*, Chapman and Hall, New York, 1982.
20. Prasad, R., Verma, S. K., Agrawal, G. G. and Mathur, N., Prediction model for peak expiratory flow in North Indian population. *Indian J. Chest Dis. Allied Sci.*, 2006, **48**, 103–106.
21. Halbert, W. A., A heteroskaedasticity – consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 1980, **48**, 817–838.
22. Friel, C. M., Weighted Least Square Regression, Criminal Justice Center, Sam Huston State University, 2006; http://www.shsu.edu/~icc_cmf/cj_789/weightedLeastSquares2.doc
23. Godfrey, K., Simple linear regression in medical research. *N. Engl. J. Med.*, 1985, **331**, 1629–1636.
24. Hackney, O. J. and Hocking, R. R., Diagnostic techniques for identifying data problem in multiple linear regression. In Proceedings of the Statistical Computing Section in Multiple Linear Regression, American Statistics Association, 1979, pp. 94–98.
25. Belsey, D. A., Kuh, E. and Welsch, R. E., *Regression Diagnostics: Identifying Influential Data and Source of Collinearity*, John Wiley, New York, 1980.
26. Khosla, T., Indices of ventilatory measurements. *Br. J. Prev. Soc. Med.*, 1971, **25**, 203–209.
27. Cole, T. J., The influence of height on the decline in ventilatory function. *Int. J. Epidemiol.*, 1974, **3**, 145–152.
28. Milledge, J. B., Vital capacity and forced expiratory volume in South Indian men. *Indian J. Chest Dis.*, 1965, **7**, 95–105.
29. Cotes, J. E., *Lung Function Assessment and Application in Medicine*, Blackwell Scientific, Oxford, 1975.

Received 16 November 2006; revised accepted 30 July 2007

Estimation of effect of gasoline quality improvement on reduction of air toxic emissions in Dhaka

Rajesh B. Biniwale[1,*] and A. F. Balloffet[2]

[1]National Environmental Engineering Research Institute, Nagpur 440 020, India
[2]Balloffet International, LLC, 2306 Eastwood Drive, Fort Collins, CO 80525, USA

Gasoline quality improvement alone can reduce total pollution load from vehicles to a considerable extent. A spreadsheet-based model has been developed in the present study to demonstrate this. An annual emission inventory for gasoline-driven vehicles was prepared with respect to volatile organic compounds, toxic air pollutants and nitrogen oxides using vehicle population data, growth rate of vehicles and usage characteristics in Dhaka. Results of the present and projected air toxic emission inventory using planned improvement in gasoline quality reveal that an emission reduction of about 20% may be achieved. A significant reduction of about 60% in pollutants like benzene justifies gasoline quality improvement in a developing country like Bangladesh.

Keywords: Exhaust volatile organic compounds, gasoline quality, polycyclic organic matter, toxic air pollutants.

MAJOR studies on automotive fuel and its impact on air quality include European Programme on Emissions, Fuel and Engine Technologies (EPEFE), Air Quality Improvement Research Program (AQIRP) of USA (also known as auto-oil) and Japan's Clean Air Programme (JCAP). The general conclusions drawn from these studies are that effects on emissions, of fuel quality changes alone, are not as significant as changes in engine technology. However, fuel quality and vehicle technologies are linked intimately and have a combined effect on emission levels[1]. This is evident from experiences of various countries that have phased out lead in gasoline and introduction of catalytic converters for reducing exhaust emissions. To phase out lead, while maintaining the octane number, there is a requirement to increase aromatics or to use oxygenates or for further severe refining. Increase in aromatics in gasoline typically results in increased benzene emissions from automotive exhaust. The key components of gasoline having an impact on air quality are lead content, sulphur content, Reid vapour pressure and benzene content. The total toxic emission from a vehicle largely depends on gasoline quality.

In Bangladesh all vehicles are imported, and there is little control over changes in engine or vehicle technology. A solution could be alternative sources of vehicles with improved technology. However, this becomes difficult as most of imported passenger cars are used vehicles. Therefore, improvement in air quality may be achieved by improving fuel quality. In view of these considerations, strategies for improvement of gasoline quality in Bangladesh, particularly in the city of Dhaka as a case study, are discussed in this communication. The effect of gasoline quality on total emission load of air toxics has been estimated by a fuel quality model. Gasoline quality parameters for the future have been suggested using the guidelines from the World Wide Fuel Charter (WWFC).

There have been various modelling studies for the estimation of pollution load from vehicles[2–4]. These studies estimated carbon monoxide (CO), hydrocarbons (HC), oxides of nitrogen (NO_x) and particulate matter (PM) emission loads. Very few of the earlier studies have estimated pollution load for air toxics such as volatile organic compounds (VOCs). The US Environmental Protection Agency (USEPA) has developed a model to calculate emissions of VOCs, toxic air pollutants (TAP), and NO_x from

*For correspondence. (e-mail: rb_biniwale@neeri.res.in)