



Location of Residences of Study Subjects

Appendix 1: Measurement Error Corrections

A. Error due to estimation of PM_{10} from TSP:

Prior to 1987, directly measured PM_{10} data were not available. PM_{10} data were estimated from measured TSP data.

Suppose that we have a response variable Y (measure of lung function) which is distributed normally. We are interested in the estimation of the slope of measured PM_{10} in the following simple linear model (i is subject index):

$$y_i = \beta_0 + \beta_1 PM10_{measured_i} + e_i \quad (1)$$

It is reasonable to assume that the estimated PM_{10} and the “true” measured PM_{10} following an additive measurement error model:

$$PM10_{estimated_i} = PM10_{measured_i} + d_i \quad (2)$$

We also further assume the following:

-independence among $e_i, PM10_{measured_i}, d_i$

- $e_i \sim N(0, \sigma_e^2)$ and $d_i \sim N(0, \sigma_d^2)$

- $PM10_{measured_i} \sim N(\mu_m, \sigma_m^2)$

Therefore, $PM10_{estimated_i} \sim N(\mu_{ex}, \sigma_{ex}^2)$, and $\sigma_{ex}^2 = \sigma_m^2 + \sigma_d^2$, and S_{ex}^2, S_m^2, S_d^2 are the corresponding sample variance estimates.

From (1), we have

$$E(y_i / PM10_{estimated_i}) = E(\beta_0 + \beta_1 PM10_{measured_i} + e_i / PM10_{estimated_i})$$

$$E(y_i / PM10_{estimated_i}) = \beta_0 + \beta_1 E(PM10_{measured_i} / PM10_{estimated_i})$$

$$E(y_i / PM10_{estimated_i}) = \beta_{b0} + \beta_{b1} PM10_{estimated_i} \quad (3)$$

Thus the β_{b0}, β_{b1} are biased estimates of the β_0 , and β_1 .

Once the β_{b0}, β_{b1} are obtained from (3), the adjustment is made as follows (Carroll et al., 1995):

$$\left\{ \begin{array}{l} \beta_0 = \beta_{b0} - (1 - \frac{S_{ex}^2 - S_d^2}{S_{ex}^2}) \beta_{b1} * \overline{PM10_{estimated}}, \quad \text{where } \beta_1 = \beta_{b1} * \frac{S_{ex}^2}{S_{ex}^2 - S_d^2} \\ \text{and } \overline{PM10_{estimated}} \text{ is the sample mean of } \mu_{ex} \end{array} \right. \quad (4)$$

The only quantity that needs to be computed from the “calibrated” sample is S_d^2 . One can simply obtain an estimate of this quantity by taking the sample variance of the difference ($PM10_{estimated} - PM10_{measured}$). The other quantities are readily available from the current sample data. The standard errors of β_0 , and β_1 can also be easily computed from (4) from the variance and covariance of β_{b0}, β_{b1} in (3).

B. Measurement error of PM10 and adjustment in multiple linear regression:

It is very likely that there are covariates which need to be controlled for in the regression model which includes PM10. We assume here that the covariates are measured without measurement errors and that they may or may not be associated with the “true” measured PM₁₀ variable.

In the case where there is independence between the “true” measured PM₁₀ and these covariates, then the adjustment can be done as in the case of the simple linear model described above. Otherwise, when there is an association, then the bias occurs not just for the estimated PM₁₀ but also for the coefficients of the other covariates (e.g., ozone). Thus, in this case there are two regression coefficients of interest (PM10, and OZONE) that are biased (β_{b1}, β_{b2}) by using the estimated PM₁₀ variable. The adjustment can be made as follows.

$$\begin{aligned}
 E(y_i / PM10_{estimated_i}) &= E(\beta_0 + \beta_1 PM10_{measured_i} + \beta_2 OZONE_i + e_i / PM10_{estimated_i}) \\
 E(y_i / PM10_{estimated_i}) &= \beta_0 + \beta_1 E(PM10_{measured_i} / PM10_{estimated_i}) + \\
 &\quad \beta_2 E(OZONE_i / PM10_{estimated_i}) \\
 E(y_i / PM10_{estimated_i}) &= \beta_{b0} + \beta_{b1} PM10_{estimated_i} + \beta_{b2} OZONE_i
 \end{aligned} \tag{5}$$

First, using the “calibrated” or past data, as in this case, 1988 to 1991 data, perform the following regression:

$$PM10_{measured_i} = \alpha_0 + \alpha_1 OZONE_i \tag{6}$$

Obtain the mean square error from this regression model, called S_α^2

$$\beta_1 = \frac{S_\alpha^2 + S_d^2}{S_\alpha^2} \beta_{b1} \text{ and } \beta_2 = \beta_{b2} - (1 - \frac{S_\alpha^2}{S_\alpha^2 + S_d^2}) \beta_{b1} \alpha_1 \tag{7}$$

It is quite often that only β_1 is of interest here, and there may be more than one measurement-error-free covariates such as WEIGHT, that could be correlated with the true measured PM₁₀. Then in this case, WEIGHT is entered into (6) and the estimated mean square error term can be obtained to use in (7).

The corrected variance of the corrected coefficient of O3 would just be the product of the variance of the uncorrected coefficient of O3 which you have from the SE of the output of the regression model and the square of the correction factor (which is the ratio of the sample variance terms from the equation 4 in my document).

C. Correction of Measurement Error where multiple variables have measurement error

1) $FEF_{75} = \text{covariates} + O_3 + O_3 * FEF_{25-75} / FVC$

For simplification, but without loss of generality, let us assume that covariates here would just be WEIGHT. We can write model 1 as:

$$E(FEF_{75}) = \beta_0 + \beta_1 O_3 + \beta_2 O_3 * FEF_{25-75} / FVC + \beta_3 WEIGHT \quad (1)$$

O_3 and $O_3 * FEF_{25-75}$ have measurement errors.

Let W be the matrix of dimension $n \times 2$ of these two measured variables. Let X be the corresponding matrix of the true values of O_3 and $O_3 * FEF_{25-75}$. Let Z be the $n \times 2$ matrix of the fixed covariates: the intercept 1's, and WEIGHT. Under the assumption of classical measurement error, we state that:

$$W = X + U, \text{ where } U \sim \text{multi variate Normal}(0, \Sigma_{uu}) \quad (2)$$

Σ_{uu} is the “known” variance-covariance matrix of the measurement errors obtained from “validated” or past data. It has a dimension of 2×2 in this case.

The corresponding sample covariance matrices of W and X are Σ_{ww} and Σ_{xx} . From (2), we also have $\Sigma_{xx} = \Sigma_{ww} - \Sigma_{uu}$. We have Σ_{ww} estimated from the sample, and Σ_{uu} estimated or known from past data or validated sample.

Let the vector $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ be the true vector of parameters to be estimated taking into account measurement error of O_3 and $O_3 * FEF_{25-75}$ (we assume that the covariate AGE is measured without error), and let the vector $\beta^* = (\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*)^T$ be the vector of the estimated coefficients obtained with measurement errors are not taken into account. Then decompose the vector β into two parts called $\beta_x = (\beta_1, \beta_2)^T$ which contains the parameters with measurement errors. The second part is $\beta_z = (\beta_0, \beta_3)^T$ which contains the intercept and the fixed, no-measurement-error covariate AGE. We are interested in obtaining estimates for $\beta_x = (\beta_1, \beta_2)^T$. The decomposition is applied to β^* as well giving $\beta_x^* = (\beta_1^*, \beta_2^*)^T$ and $\beta_z^* = (\beta_0^*, \beta_3^*)^T$. The solution for $\beta_x = (\beta_1, \beta_2)^T$ is as follows:

$$\begin{pmatrix} \beta_x \\ \beta_z \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{xx} + \Sigma_{uu} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix} \begin{pmatrix} \beta_x^* \\ \beta_z^* \end{pmatrix} \quad (3)$$

for (3), we have $\begin{pmatrix} \beta_x^* \\ \beta_z^* \end{pmatrix}$ from running model (1) without taking into account any measurement errors. We can estimate $\Sigma_{xx} = \Sigma_{ww} - \Sigma_{uu}$. We can also estimate Σ_{xz} by using “validated data set” or by using Σ_{wz} (2x2 matrix). We have Σ_{zz} , sample covariance matrix from the intercept vector 1’s and AGE. $\Sigma_{xx} + \Sigma_{uu}$ is equal to the sample Σ_{ww} . So all the quantities in (3) are computable.

Let (3) be written as $B = VB^*$, where $B = \begin{pmatrix} \beta_x \\ \beta_z \end{pmatrix}$, and $B^* = \begin{pmatrix} \beta_x^* \\ \beta_z^* \end{pmatrix}$, and

$$V = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{xx} + \Sigma_{uu} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix}, \text{ then Variance of } B \text{ is } V^T \text{var}(B^*)V \quad (4)$$

where $\text{var}(B^*)$ is obtained from running model (1).

For the model:

$$FEF_{75} = \text{covariates} + O_3 + O_3 * FEF_{2575} / FVC + PM_{10}$$

or

$$E(FEF_{75}) = \beta_0 + \beta_1 O_3 + \beta_2 O_3 * FEF_{25-75} / FVC + \beta_3 PM_{10} + \beta_4 WEIGHT$$

the idea is similar to the previous case but with the vector of parameters $\beta_x = (\beta_1, \beta_2, \beta_3)^T$.

The construction of the matrix components used in (3) would be similar for this case as well.

Reference:

Carroll RJ, Ruppert D, Stefanski LA (1995). Measurement Error in Nonlinear Models. London: Chapman & Hall.