# Effects of Air Pollution and Lung Function in Children with Asthma
## By: Thuy Nguyen

## Introduction:

There exist concerns about how chronic exposure to highly polluted air may affect lung function and development in children with sensitive airways. A 10-year prospective study, starting in 1993, has been annually evaluating lung function based on the forced expiratory volume in one second measured in ml for children of 4th and 7th-grade cohorts. A subset of the data includes children with asthma from two groups: one lives in an area with an average annual $NO_2$ greater than 30.0ppb, and the other lives in an area with an average yearly $NO_2$ of less than 30.0ppb. We will use this subset to assess whether air pollution (measured by $NO_2$ concentration) has a detrimental effect on lung function (measured by FEV1) in children with asthma.

## Method

*Descriptive Analysis:* We present exploratory data analysis that investigates the scientific questions using figures and summary statistics. We present the baseline characteristics of participants by the annual $NO_2$ concentration in the left panel of Table 1 and a summary of the Distribution of log(FEV) at different timepoints in Table1 right. To investigate the relationship between FEV1 and time, we plot FEV1 as a function of age and height in Figure 1(a) and (b), respectively. In Figure 1(c)–(d), we display average log FEV1 as a function of age and height, stratified by the annual $NO_2$ concentration of the study subjects' residency. We also plot the pairwise correlations between height, age, and log(FEV1) in Figure 2. To visualize the missing data, we plot a histogram of missing data and the missing pattern in Figure 3.

*Confirmatory Analysis:* Recall that we wanted to know if children living in areas with a high level of pollution appear to have a poorer level of lung function; and whether pollution is associated with poor development or growth in lung function. Previous research provides evidence for an approximately linear relationship between log(FEV1) and age in children. Therefore, we focus on the mean model of log FEV as a function of indicator of $NO_2$ levels, sex, age, and height, and the interaction of age and $NO_2$ levels. Since the measures within a cluster are highly correlated, the choice of fitting model is GEE with an independent working correlation matrix. Any options of working correlation matrix led to valid inference asymptotically for the GEE model. Before fitting the model, however, we will perform multiple

imputations in the long format. We construct Wald-based 95% confidence intervals using the output.

**Results:**

*Descriptive Analysis:* Overall, baseline age, height, sex, and the number of follow-up visits are roughly equal between the areas with a high N02 concentration and those with a low N02 concentration. In addition, children enrolled in the study were followed approximately annually. Therefore, FEV1 measurements are roughly grouped around years 1,2,3 and 4 post baseline (Table 1 right). Overall, children with asthma living in areas with high N02 have a slightly lower log (FEV) than their counterparts living in the areas with low N02.

Figure 1 (a)-(b) show that the average FEV increases as a function of both age and height nonlinearly. There is evidence of lower log FEV1 for children with asthma who live in areas with high N02, especially when children are older than 14 or taller than 170 cm (Figure 1 (c)-(d)). Also, the rate of change in log FEV1 seems to be constant for height, but non-constant for age. The plots of pairwise correlations between height, age, and log (FEV1) in Figure 2 show that subject in areas with higher NO2 result in slightly lower correlations. The plot of missing pattern (Figure 3 (a) right) indicates that 28% of people are completely observed, 15.8% missing their second and third responses and so on. The Figure 3 (a) left show a histogram of missing data by column in our dataset. About 70% of responses at the last follow-up (fev1.5) are missing. Figure 3 (b) shows no pattern of missing data between areas of low versus high N02 concentration. In this figure, blue means missing; yellow means observed.

*Confirmatory Analysis:* We find that the rate of change of log FEV1 per year for children with high NO2 exposure is 0.01 (95% CI: 0.0001–0.0206, p-value: 0.047) lower than those with low NO2 exposure, after adjusting for height and sex. In addition, at 5% significance level, we have no evidence that children with asthma living in areas with high levels of pollution have poorer lung function when compared to those in areas with low levels of pollution, after adjusting for height and sex (p-value: 0.14).

**Conclusion:**

There is no association between living in a high concentration of N02 areas and lung capacity function in children with asthma after adjusting for height, age and sex. However, pollution is associated with a poor development or growth in lung function.

# Tables and Figure

| | Low NO2 (N=214) | High NO2 (N=153) | Overall (N=367) | | Low NO2 (N=214) | High NO2 (N=153) | Overall (N=367) |
|---|---|---|---|---|---|---|---|
| Baseline Age (Year) | | | | Baseline | | | |
| Mean (SD) | 11.1 (1.44) | 11.0 (1.55) | 11.1 (1.48) | Mean (SD) | 7.73 (0.231) | 7.72 (0.227) | 7.72 (0.229) |
| Median [Min, Max] | 10.4 [8.73, 15.4] | 10.2 [8.50, 14.4] | 10.3 [8.50, 15.4] | Median [Min, Max] | 7.72 [7.07, 8.41] | 7.69 [7.29, 8.39] | 7.71 [7.07, 8.41] |
| Baseline Height (Cm) | | | | 1 Year | | | |
| Mean (SD) | 146 (10.5) | 147 (11.5) | 147 (10.9) | Mean (SD) | 7.83 (0.231) | 7.81 (0.247) | 7.82 (0.238) |
| Median [Min, Max] | 145 [124, 179] | 145 [124, 177] | 145 [124, 179] | Median [Min, Max] | 7.81 [7.21, 8.52] | 7.79 [7.21, 8.43] | 7.81 [7.21, 8.52] |
| Sex | | | | Missing | 2 (0.9%) | 1 (0.7%) | 3 (0.8%) |
| Female | 92 (43.0%) | 62 (40.5%) | 154 (42.0%) | 2 Year | | | |
| Male | 122 (57.0%) | 91 (59.5%) | 213 (58.0%) | Mean (SD) | 7.96 (0.237) | 7.91 (0.228) | 7.94 (0.234) |
| No of Follow-ups | | | | Median [Min, Max] | 7.95 [7.29, 8.51] | 7.88 [7.42, 8.48] | 7.92 [7.29, 8.51] |
| 1 | 2 (0.9%) | 1 (0.7%) | 3 (0.8%) | Missing | 28 (13.1%) | 24 (15.7%) | 52 (14.2%) |
| 2 | 26 (12.1%) | 23 (15.0%) | 49 (13.4%) | 3 Year | | | |
| 3 | 38 (17.8%) | 20 (13.1%) | 58 (15.8%) | Mean (SD) | 8.07 (0.219) | 8.00 (0.228) | 8.04 (0.225) |
| 4 | 93 (43.5%) | 61 (39.9%) | 154 (42.0%) | Median [Min, Max] | 8.07 [7.42, 8.67] | 7.99 [7.45, 8.54] | 8.04 [7.42, 8.67] |
| 5 | 55 (25.7%) | 48 (31.4%) | 103 (28.1%) | Missing | 66 (30.8%) | 44 (28.8%) | 110 (30.0%) |
| | | | | 4 Year | | | |
| | | | | Mean (SD) | 8.14 (0.192) | 8.05 (0.185) | 8.10 (0.193) |
| | | | | Median [Min, Max] | 8.11 [7.57, 8.58] | 8.06 [7.56, 8.46] | 8.09 [7.56, 8.58] |
| | | | | Missing | 159 (74.3%) | 105 (68.6%) | 264 (71.9%) |

Table 1: *Left* : "Distribution of baseline covariates and number of follow-up"

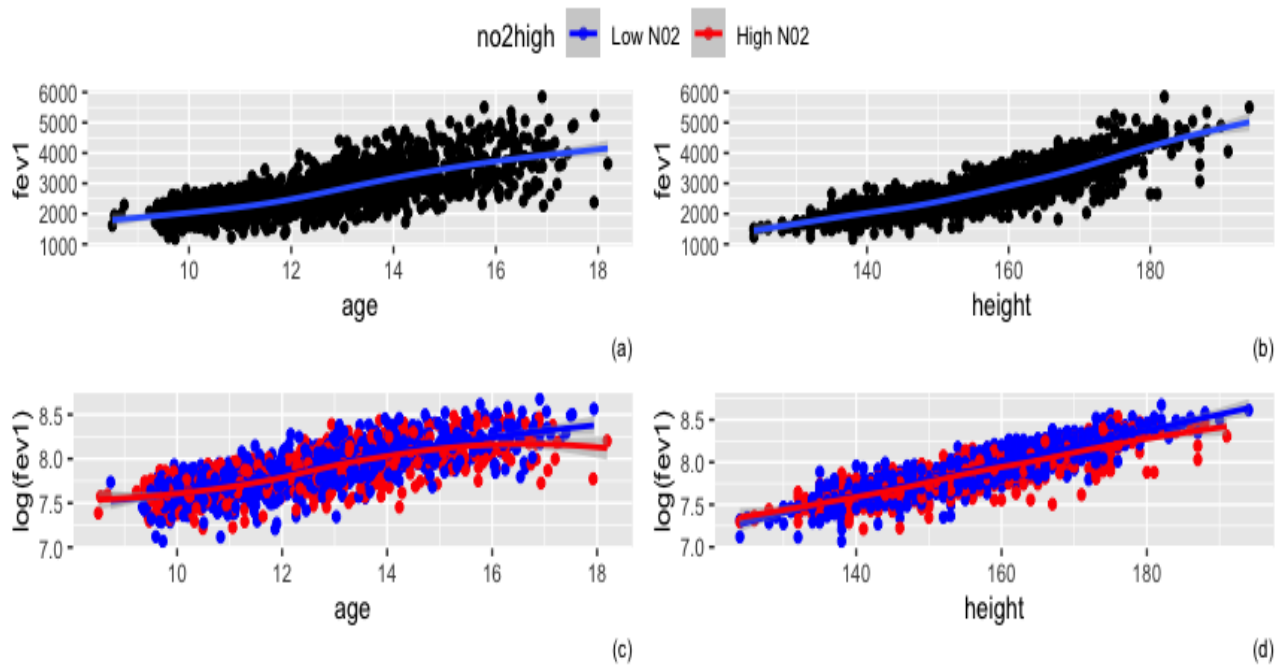   *Right:* "Summary of distribution of log(FEV) at different timepoints")

Figure 1: (a): FEV1 over age with LOESS smoothing curve. (b): Same as (a), but for FEV1 over height. (c): log FEV over age, stratified by average annual NO2 concentration. (d): log FEV over height, stratified by average annual NO2 concentration.
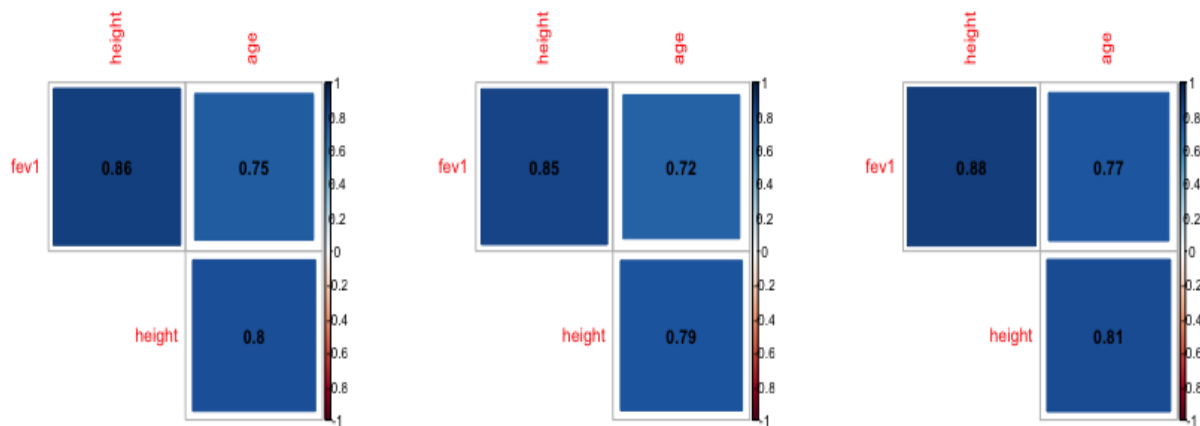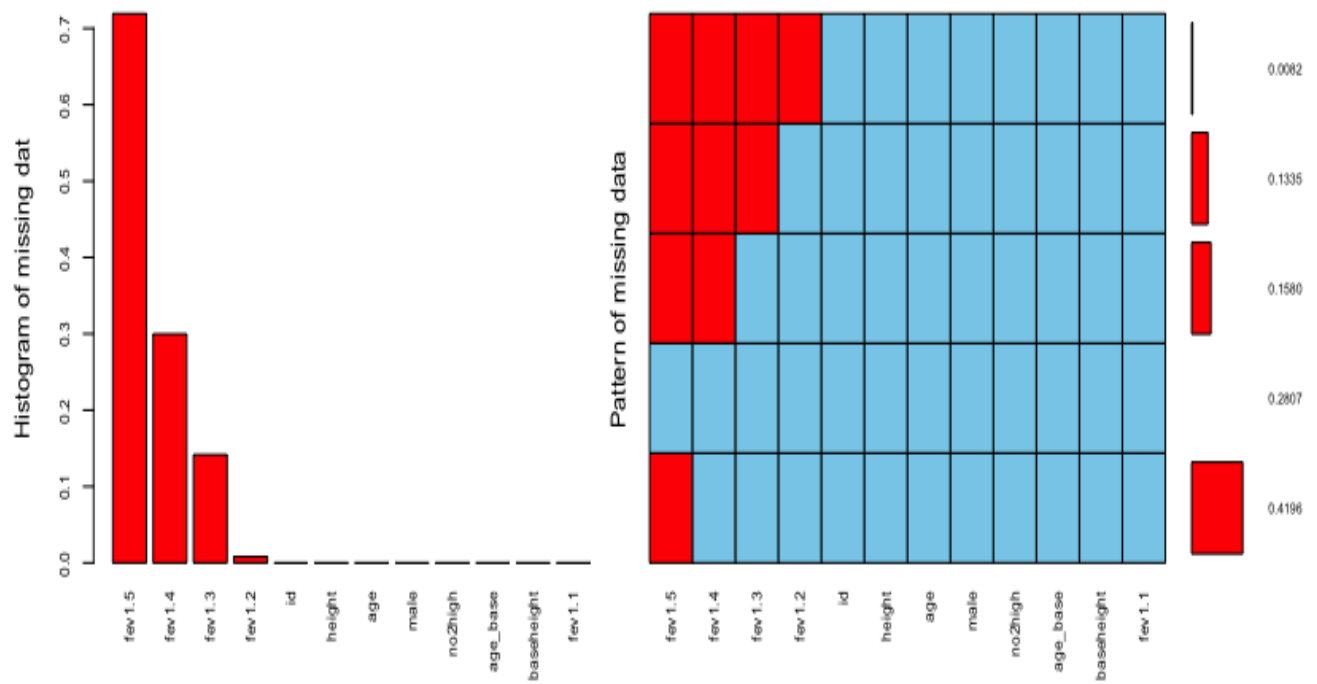
Figure 2: (a): Pairwise correlations among height, age, and log(FEV1) based on all available observations. (b): Same as (a), but for subjects in the high NO2 concentration area. (c): Same as (a), but for subjects in the low NO2 concentration area.
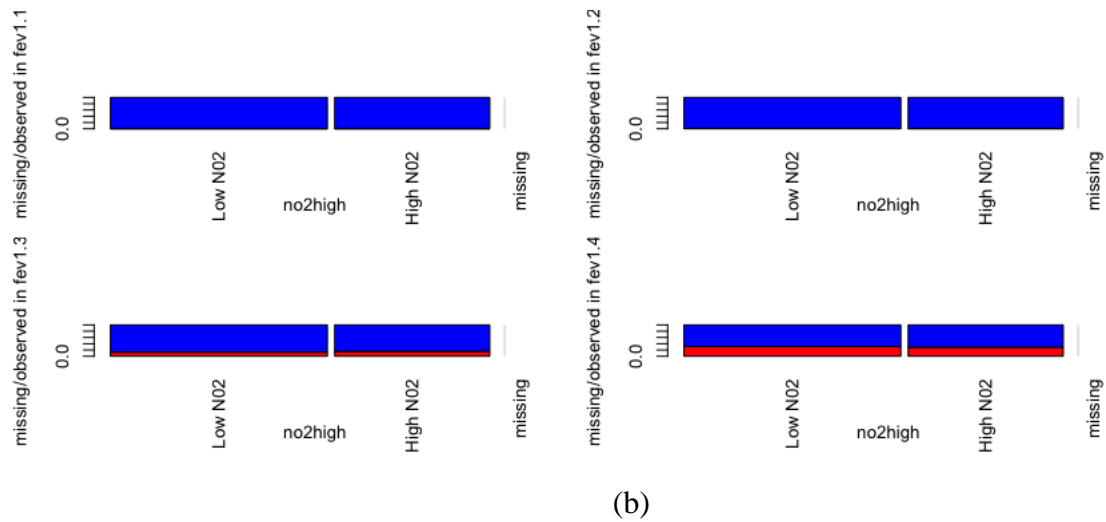


(a)

(b)

Figure 3: (a): Missingness pattern of outcomes

(b) Missing pattern of outcomes by follow-up visits, stratified by N02 levels