Report

<u>Introduction:</u> 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 NO2 greater than 30.0ppb, and the other lives in an area with an average yearly NO2 of less than 30.0ppb. We will use this subset to address two questions: "Do children living in areas with a high level of pollution have a poorer level of lung function? And "Is pollution associated with poor growth in lung function?"

<u>Method:</u> *<u>Descriptive Analysis</u>*: Height and age are highly correlated in children in grades 4th and 7th. So, we will investigate the effect of high N02 on lung function and development under the impact of age and height separately. We present a mean and standard deviation of fev1 by NO2 levels and equally divided age intervals in table 1; fev1 by NO2 levels, and equally divided height intervals in table 2. Also, we present scatterplots of fev versus time by N02 levels groups (Figure 1) and fev versus height by N02 levels groups (Figure 2).

<u>Confirmatory Analysis</u>: To determine the effect of N02 levels on lung function and growth in the lung function of children with asthma, we will use an autoregressive correlation matrix and analyze the response profiles: we include N02 levels, age-since-entry, and their interaction. The alternative model uses N02 stories, height-since-entry, and their interaction. These allow for an arbitrary pattern in mean fev over time and different relationships between the two groups. The measurements are made at approximately equal intervals of time. Also, individuals are in grades 4 and 7; biologically, their fev quickly increases every year, so choosing the autoregressive correlation model is reasonable. With highly correlated outcomes and the small size of the clusters, we use the GEE approach to obtain point estimates and standard error for the regression parameters. We use a log link function for function fev with heigh since the growth of fev (Figure 2) appears to be exponential, and identity link function for function with age. We will compare the mean fev at baseline of children in a high N02 area differs from the

mean fev of children in a low N02 area and compare the rate of changes in mean fev of the two groups. We will use a Wald test to if the differences are statistically significant at the 5% level.

Results: *Descriptive Analysis:* The data has 264 individuals who missed at least one annual measurement. Figure 1 and Figure 2 visually show that the fev tends to increase as age and height increase. While fev grows about linearly over the age, it grows exponentially over the height. There are considerable overlaps between the mean fev of the two groups in both figures. We only see a significant difference in average fev of the two groups when age is above 15 (Figure 1) and height is over 165 cm (Figure 2). The group with low N02 seems to have a slightly faster increase in fev over age and height than the other. Also, Table 1 and Table 2 confirm the visualizations. Table 1 and Table 2 numerically represent that the variance of fev increases as age and height increase.

Confirmatory Analysis: There is no statistically significant evidence at the 5% level that at baseline, children with asthma who live in areas with high N02 levels have worse lung function than those who live in areas with low N02 in both models. Also, there is no statistically significant evidence at the 5% level that children with asthma who live in areas with high levels of N02 have poorer growth in lung function for every 1 cm difference in height than those who live in areas with a low of N02 level. However, we find statistically significant evidence that for every one-year difference in age, children with asthma who live in the areas with high levels of N02 have worse growth in lung function than those who live in the areas with low N02 levels. The point estimate and 95% confidence interval are presented in Table 3. Compared to children with asthma living in areas with low N02 levels, those living in areas with high N02 levels have an additional 36.8ml decrease (95% CI: --66.3 ml – -7.33 ml) in mean fev for every one-year difference from baseline.

Tables:

2321 (559)

Age	High NO2	Low NO2	All
[8.5-10.5]	2023 (286)	2067 (320)	2045 (304)
(10.5-12.5]	2336 (373)	2339 (425)	2338 (405)
(12.5-14.5]	2942 (578)	3004 (525)	2978 (548)
(14.5-16.5]	3480 (720)	3638 (671)	3576 (693)
[16.5-18.5]	3503 (735)	4113 (863)	3854 (857)

Table1: Mean (and standard deviations) of fev1 (forced expiratory volume in 1 second) (ml) by NO2 levels and age (years)

Height	High NO2	Low NO2	All
[124-138]	1860 (232)	1825 (286)	1840 (264)
(138 -152]	2201 (302)	2257 (278)	2232 (290)
(152-166]	2781 (390)	2920 (418)	2865 (412)
(166 -180]	3579 (600)	3689 (543)	3640 (571)
(180 -194]	4099 (659)	4735 (457)	4504 (613)

Table2: Mean (and standard deviation) of fev1 (ml) by NO2 levels and height (cm)

	Low N02	High N02	Difference
Rate of growth From baseline	300.3 (280.8 - 319.9)	263.5 (241.4 – 285.6)	-36.8 (-66.3 – -7.33)

Table 3: Change of mean Fev (ml) from baseline

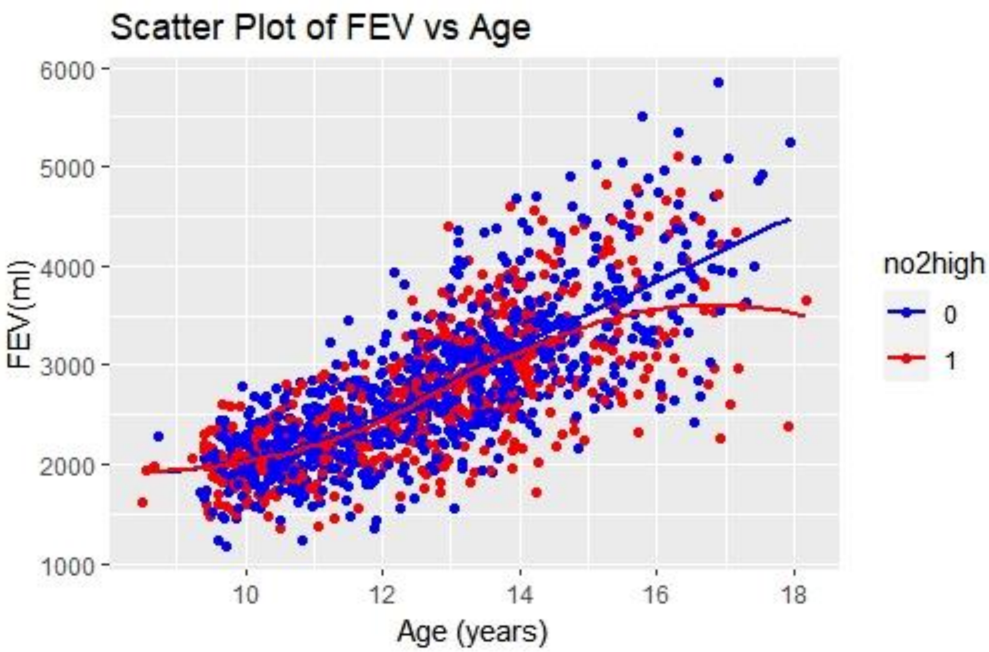

Figure 1: Scatterplot of Fev(ml) vs Age (years)

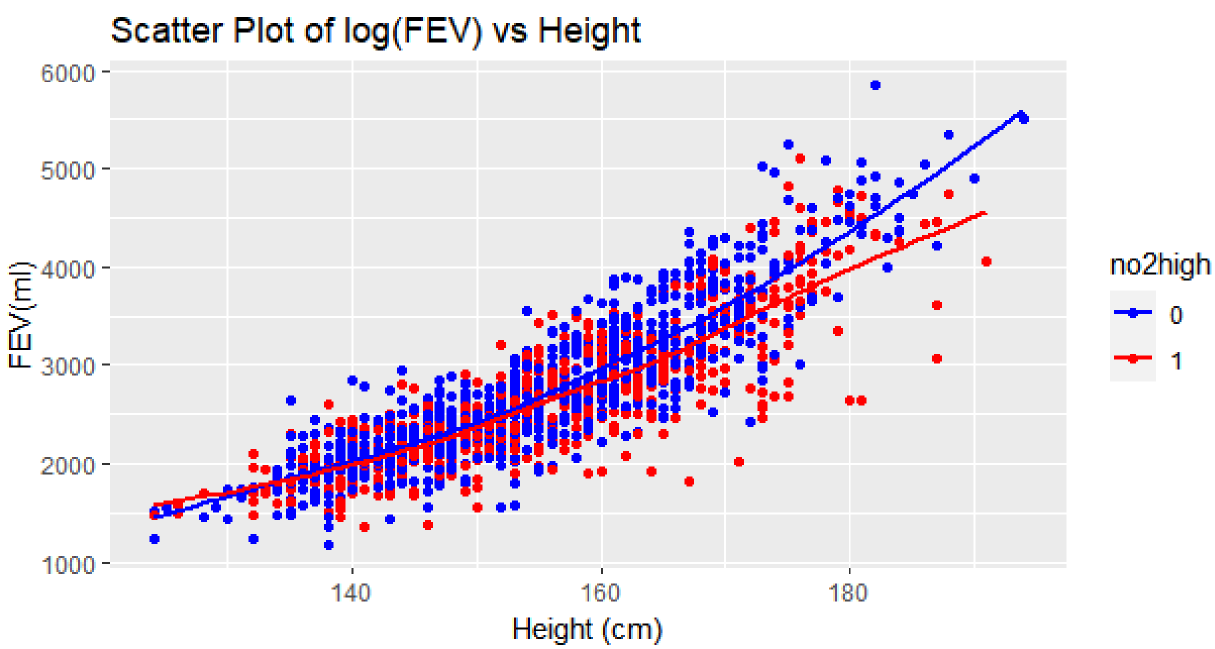

Figure 2: Scatterplot of Mean Fev (ml) vs Height (cm)

CODE:

library(reshape2)

library(ggplot2)

library(tidyr)

library(dplyr)

library(corrplot)

library(nlme)

library(geepack)

library(multcomp)

Load data

asth<- read.csv("AsthmaNO2.csv")

asth <- asth[, !(names(asth) %in% c("X","asthma"))]

asth$no2high <- as.factor(asth$no2high)

Create age at baseline

asth$age_base <- rep(0, nrow(asth))

for (i in unique(asth$id)){

  asth$age_base[asth$id ==i]<- min(asth$age[asth$id==i])

}

Create age since entry

asth$dage <- asth$age - asth$age_base

create height at baseline

```r
asth$baseheight <- rep(0, nrow(asth))

for (i in unique(asth$id)){

  asth$baseheight[asth$id ==i]<- min(asth$height[asth$id==i])

}

# Create height since entry

 asth$dheight <- asth$height - asth$baseheight


# Exploratory data

# Missing: yes

ast_wide <- reshape(asth[, c("id", "fev1", "time")], direction = "wide", id = "time", idvar = "id")

nrow(ast_wide[!complete.cases(ast_wide),])


# Summary data

asth$age_cut <- cut(asth$age, breaks = c(8.5,10.5,12.5,14.5,16.5,18.5), include.lowest = T)

asth$height_cut <- cut(asth$height, breaks = c(124,138,152,166,180,194), include.lowest = T)

asth %>%  group_by(no2high, age_cut) %>% summarise(fev_h = mean(fev1), sd_h = sd(fev1))

asth %>%  group_by(no2high, height_cut) %>% summarise(fev_h = mean(fev1), sd_h = sd(fev1))

asth %>%  group_by( age_cut) %>% summarise(fev_h = mean(fev1), sd_h = sd(fev1))

asth %>%  group_by(height_cut) %>% summarise(fev_h = mean(fev1), sd_h = sd(fev1))

# Scatterplots

color <- c("blue", "red")

par(mfrow=c(1,2))
```

ggplot(data =asth, aes((age),(fev1),col = no2high))+geom_point() + geom_smooth() + scale_color_manual(values = color) + ylab("FEV(ml)") + xlab("Age (years)") +ggtitle("Scatter plot of FEV vs Age")

ggplot(data =asth, aes((height), (fev1),col = no2high))+geom_point() + geom_smooth(se = F)+ scale_color_manual(values = color) + ylab("FEV(ml)") + xlab("Height (cm)") + ggtitle("Scatter Plot of log(FEV) vs Height")

 # GEE analyses

asth<- asth[order(asth$id, asth$age),]

mod1 <-geeglm((fev1)~ (dage)*no2high, data= asth, id = id, corstr = "ar1")

summary(mod1)

anova(mod1)

tidy(mod1, conf.int = T)

mod2 <-geeglm((fev1)~ (dheight)*no2high, data= asth, id = id, corstr = "ar1")

summary(mod2)

anova(mod2)