BIOST 540 Midterm

Katie McFarlane

Introduction

Forced expiratory volume (FEV) is a measure of lung function that generally increases as children age and their lungs continue to develop. Asthma can be diagnosed using FEV results. As such, children with asthma will typically have lower FEV1 values (FEV in 1 second measured in ml) than children who are not asthmatic. A longitudinal study of lung function was conducted in 367 asthmatic 4^{th} and 7^{th} graders in Southern California. This data is dichotomized by pollution level (high pollution is considered an average annual NO2 of >30ppb, with low pollution an average annual NO2 <=30ppb). I am interested in analyzing the association between pollution level and FEV1 values of asthmatic children over time to investigate if there is a difference in lung function and growth depending on pollution level.

Methods

Descriptive Analysis: I present descriptive statistics (mean and standard deviation) illustrating the relationship between FEV1 and age category (Table 1) and FEV1 and height category (Table 2) for children in high and low pollution areas, as well as a plot of individual regression slopes for each participant (Figure 1). Additionally, I calculated the correlation matrix for all participants (Figure 2). Measurements were taken roughly one year apart and up to five times per individual, so measurements were rounded to the nearest full year and ordered by the number of years since baseline that the measurement was taken to perform this calculation. *Confirmatory Analysis*: I utilized a generalized estimating equations (GEE) approach with robust standard errors and an autoregressive working correlation matrix to investigate the association between pollution level and FEV1 in children with asthma, with variables for FEV1 value, pollution level, age, and an interaction term between pollution level and age. The working correlation matrix was chosen based on information in Figure 2, and a GEE approach was chosen in order to account for clustering by individual participant and to estimate valid coefficients and standard errors even if the choice of correlation matrix is inconsistent with the true correlation matrix. This approach also allowed me to investigate a population average effect rather than individual specific effects while making few assumptions. In order to test if there is evidence for a change in mean FEV1 value over time given pollution level, I used an ANOVA test. I also plotted the residuals from this model as a diagnostic to see if the assumed mean model is acceptable.

Results

Descriptive Analysis: Younger children have lower average levels of FEV1 compared to older children in both pollution level groups. While the difference in the youngest age category (aged 8 to 10) is very small between those in low and high pollution areas, all other age categories indicate a higher mean FEV1 in low pollution areas, with an increasingly large difference between older children exposed to low as compared to high pollution. Note that 16 to 19-year-olds in low pollution areas have a higher mean FEV1 value by 294 ml (Table 1). A similar relationship holds for FEV1 and height: the shortest categories of participants (120 to 135 cm) have very similar mean FEV1 across pollution levels, with those in lower pollution areas having 7 ml lower mean FEV1 values, yet for all other height categories, those in lower pollution areas have higher mean FEV1 values than those of the same height category in higher pollution areas (Table 2). While the difference in mean FEV1 does not increase consistently across all height categories like it did for all age categories, the data does indicate mean FEV1 is higher for those living in low pollution areas. This relationship can also be seen from the plot of regression lines: children living in low pollution areas have steeper slopes than those in high pollution areas, with an average slope of 289 for those in low pollution areas, as compared to 256 for high pollution areas (Figure 1). We also note a positive correlation across measurement times for individuals that lessens the farther apart measurements were taken, indicating an autoregressive relationship (Figure 2). *Confirmatory Analysis*: Using GEE to examine the average linear trend between pollution, age, and an interaction term between the two, I find that the estimated difference in mean FEV1 levels for populations of age zero in high pollution areas is 338 ml greater than those in low pollution areas (95% CI: 63, 713), but that there is an additional estimated decrease of 35.3 ml in high pollution areas for every one year increase in age (95% CI: 6.1, 64.6). Coefficients for this model can be found in Table 3. Using an ANOVA test, I find this additional difference of 35.3 ml per year is statistically significant at the 0.05 level (p=0.018), so I reject the null hypothesis that there is no evidence for a change in mean FEV1 value over time given pollution level. I find similar results when using a model that includes height rather than age and note an additional decrease in FEV1 value for those in high pollution areas of 5.59 mL for every one cm of height (95% CI: 0.85, 10.33, p=0.021). Thus, based on this data I find evidence that high pollution levels have a negative impact on lung growth and function among asthmatic children over time as compared to low pollution levels.

Tables and Figures

FEV1 Values Across Age Categories

	Low NO2 levels	High NO2 levels	Difference
Aged [8, 10]	1971 (329)	1975 (291)	-4
Aged (10, 12]	2232 (360)	2184 (326)	48
Aged (12, 14]	2869 (508)	2815 (534)	54
Aged (14, 16]	3493 (611)	3311 (655)	182
Aged (16, 19]	3979 (799)	3685 (758)	294

Table 1: Mean (and standard deviation) of FEV1 values (in ml) by pollution level and the differences between groups (average FEV1 for high pollution level subtracted from average FEV1 for low pollution level). Ages are broken into five categories, each of which covers two years, excepting the fifth group, which includes one measurement taken at an age between 18-19.

FEV1 Values Across Height Categories

	Low NO2 levels	High NO2 levels	Difference
[120, 135] cm	1756 (289)	1763 (201)	-7
(135, 150] cm	2183 (292)	2128 (289)	55
(150, 165] cm	2858 (425)	2720 (381)	138
(165, 180] cm	3631 (551)	3548 (597)	83
(180, 195] cm	4735 (457)	4099 (659)	636

Table 2: Mean (and standard deviation) of FEV1 values (in ml) by pollution level and the differences between groups (average FEV1 for high pollution level subtracted from average FEV1 for low pollution level). Height is broken into five categories, encompassing 15 cm each.

Individual Regression Slopes of FEV1 Over Time

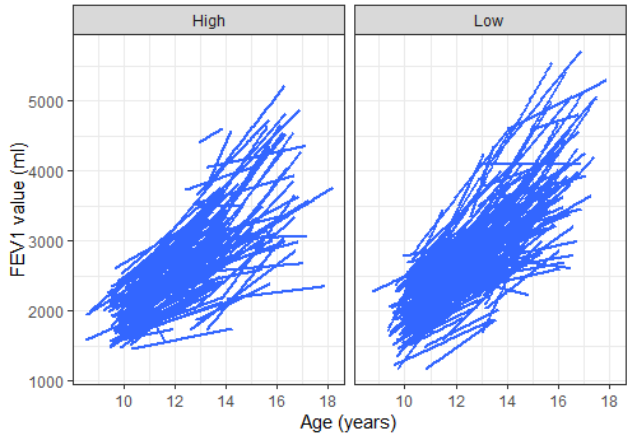

Figure 1: Individual linear regression slopes for each participant illustrating the relationship between age in years and FEV1 value in ml, separated by high and low pollution level.

Correlation between All Observations

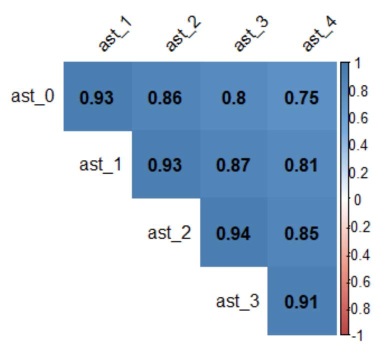

Figure 2: Correlation matrix illustrating the relationship between individuals' measurements, with ast_i indicating the participant's i^{th} measurement taken after baseline (up to 5 total).

GEE Model Coefficient Estimates Including Pollution Level and Age

	Estimate	Robust standard error	95% CI
Intercept	-1007	110	(-1223, -791)
Pollution Level	388	166	(63, 713)
Age	299	9.7	(281, 318)
Pollution Level*Age	-35.3	14.9	(-64.6, -6.1)

Table 3: Estimated coefficients, robust SE, and 95% CIs for a GEE model using an autoregressive working correlation matrix to estimate mean FEV1 values using pollution level, age, and an interaction term between pollution level and age as predictors.