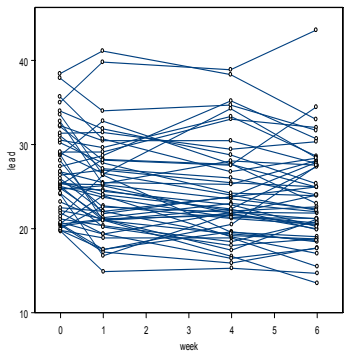
## 1. *Purpose of the analysis*

This report is a kind of preliminary analysis in order to see the treatment effect of succimer (a chelating agent) in lead-exposed children based on a randomized study with four repeated measurements of blood lead levels based on several models.

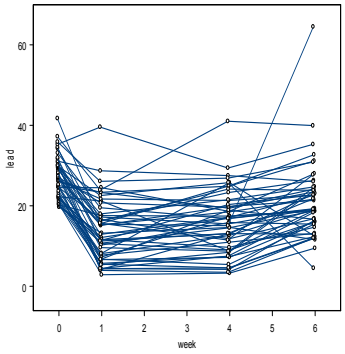
## 2. *Exploratory data analysis: Plots and Descriptive table*

1) By investigating trajectory of blood lead levels per weeks and mean plots as below, we can see that there is a different pattern between the treatment group and the placebo group. Further, a possible influential case is detected with the trajectory plot.[1]

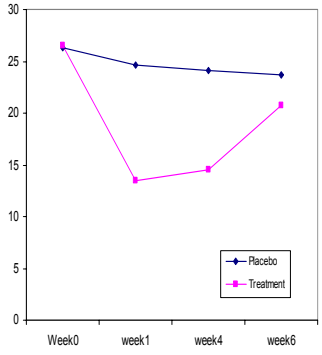
2) Based on the mean values of lead level in week 0 per each group and ANOVA test ($F=0.07$), it is found that there is no pre-existing difference in blood lead level.



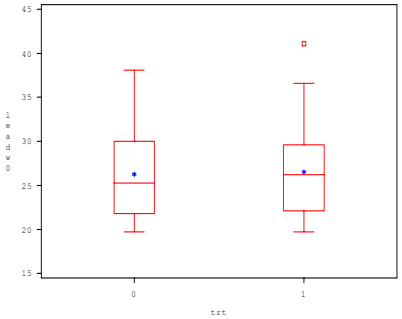

[ Placebo group]

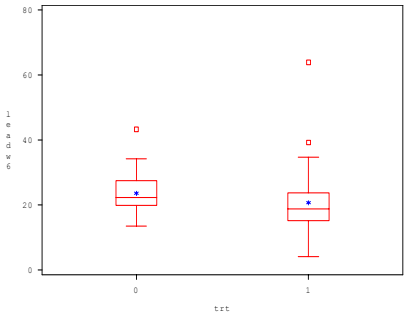



[Treatment group]




[ Mean plots per group]




| GROUP | N | Variable | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|---|
| 0 (Placebo) | 50 | leadw0 | *26.272* | 5.024 | 19.7 | 38.1 |
| | | leadw1 | 24.660 | 5.461 | 14.9 | 40.8 |
| | | leadw4 | 24.070 | 5.753 | 15.3 | 38.6 |
| | | leadw6 | 23.646 | 5.640 | 13.5 | 43.3 |
| 1 (Treatment) | 50 | leadw0 | *26.540* | 5.021 | 19.7 | 41.1 |
| | | leadw1 | 13.522 | 7.672 | 2.8 | 39.0 |
| | | leadw4 | 15.514 | 7.852 | 3.0 | 40.4 |
| | | leadw6 | 20.762 | 9.246 | 4.1 | *63.9* |

[1] This case has a high base line and the highest lead level in week 6, which is unusual compared to others (id:40 33.7, 14.9, 14.5, 63.9). Thus, we should pay attention to this data point for the parameter estimates of interest. It will be helpful to figure out the influence of possible outliers by comparing the results after deleting this case.

## *3. Statistical analysis: ANOVA, ANCOVA*

**1) Model 1 (One-Way ANOVA)**: $Y_i = \beta_0 + \beta_1 Trt_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$

With one-way ANOVA model, it is found that the overall difference of blood lead level between the two groups in week 6 ( $\hat{\beta}_1$=-2.884) is not statistically significant (p=.062).

**2) Model 2 (ANCOVA)**: $Y_i = \beta_0 + \beta_1 Trt_i + \beta_2 LeadW0_i + \varepsilon_i$

ANCOVA model was adopted in order to see the treatment effect after controlling for the initial values, in this case, in week 0. The analysis result shows that in holding the baseline value constant, children in the treatment group in week 6 tend to have 3.12 lower blood lead levels on average than those in the placebo group ( $\hat{\beta}_1$=-3.12, p=.0148). Also, the positive association between initial values and the blood lead levels in week 6 was found ( $\hat{\beta}_2$=.88, p<.001).

**3) Model 3: test for the treatment-baseline interaction**

$Y_i = \beta_0 + \beta_1 Trt_i + \beta_2 LeadW0_i + \beta_3 (Trt \times LeadW0)_i + \varepsilon_i$

Based on the model above, it is found that the estimate of treatment-baseline Interaction ( $\hat{\beta}_3$ = .06 3) was not statistically significant (p=.804).

| Source | DF | SS | MS | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 3 | 2125.33714 | 708.44571 | 17.76 | <.0001 |
| Error | 96 | 3830.40126 | 39.90001 | | |
| Corrected Total | 99 | 5955.73840 | | | |

| Root MSE : 6.31665 | R-Square: 0.3569 | Adj R-Sq: 0.3368 |
|---|---|---|

| Variable | DF | Parameter est. | SE | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 1.35389 | 4.80252 | 0.28 | 0.7786 |
| Trt | 1 | -4.79015 | 6.82742 | -0.70 | 0.4846 |
| leadw0 | 1 | 0.84851 | 0.17961 | 4.72 | <.0001 |
| INT | 1 | 0.06325 | 0.25409 | 0.25 | 0.8039 |

**4) Reanalysis for the subsets**: $Y_i = \beta_0 + \beta_1 Trt_i + \varepsilon_i$

In order to investigate whether the treatment effect differs depending on the values of baseline, the data set was divided intentionally; the first data set is for the group with the high baseline (1SD above from its mean, n=17) and the other is for the low group in terms of the baseline (1SD below from its mean, n=19). Also, the left children (n=64) were included in the third data set, which can be indicated as a middle group. This analysis can be meaningful in that we can have a sense of which group has more benefit from the treatment. The results show that for the high baseline group. there is no significant difference in blood lead level in week6

between the two groups ( $\hat{\beta}_1$=.746, p=.891)[2]. On the other hand, in terms of the group with a moderate baseline, children with the treatment tend to have 3.5 point lower on average than those with a placebo ( $\hat{\beta}_1$=-3.49, p=.007). For the low baseline group, it is found that the treatment group has 4.55 lower than placebo group ( $\hat{\beta}_1$=-4.55, p=.032).

**5) Influence of possible outliers**

The result after deleting one case (id=40) was compared to the original result above. First, with respect to one-way ANOVA, treatment effect becomes significant (( $\hat{\beta}_1$=-3.76, p=.0037) unlike the result from the original data ( $\hat{\beta}_1$=-2.884, p=.062). With the model 2 after controlling for the baseline values, there is no change for the significance of the treatment effect, but the magnitude slightly increased ( $\hat{\beta}_1$=-3.858, p=.0002). For the treatment-baseline interaction effect ( $\hat{\beta}_3$), it is not significant but the direction of the interaction effect is changed.

| Variable | DF | Parameter Est. | SE | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 1.35389 | 3.82326 | 0.35 | 0.7240 |
| trt | 1 | 0.43050 | 5.47949 | 0.08 | 0.9375 |
| leadw0 | 1 | 0.84851 | 0.14299 | 5.93 | <.0001 |
| INT | 1 | -0.16285 | 0.20450 | -0.80 | 0.4278 |

In terms of subset data, only the high baseline group with the case (if=40) was reanalyzed. The result is that even though the estimate of treatment effect is not still significant ( $\hat{\beta}_1$=-3.867, p=.28), it is noticeable that the direction and the magnitude of the estimate was changed ( $\hat{\beta}_1$=.746, p=.891 with the original data). Also, this result is consistent with it from the interaction model.

## *4. Summary and conclusion*

The final results based on this preliminary analysis can be summarized as follows.
First, it turns out that there is no pre-existing difference in blood lead level, so that we can say the randomization worked appropriately. Second, based on the interaction model, no significant treatment-baseline interaction was found, which means the treatment effect would be the same across the blood lead levels of week 0 (the assumption of parallel regression lines held). Furthermore, by rerunning one-way ANOVA with the subset models, the mean differences on blood lead level in week 6 between the treatment group and the placebo group were investigated. Even though the treatment effect for the high baseline group is no longer significant and the direction is positive, it turns out that the result is mainly due to an influential outlier. Hence, it will be helpful to investigate this unusual case in detail and keep in mind the influence of this outlier on parameter estimates of our interest. Finally based on the several models, we can conclude that there is a statistically significant treatment effect on blood lead level of children. Children who had a treatment tend to have a lower blood lead level. However, further analysis such as the rate of change of blood lead level over time would be necessary to investigate the treatment effect.

[2] However, we should be careful in interpreting the result since the possible outlier in this data set can influence on the parameter estimate.