Introduction: CD4 count has been considered an important biomarker in the HIV virus progression and AIDS diagnosis. The AIDS Clinical Trial Group (ACTG) study 193A was a double-blinded, randomized study about how the four different treatments, zidovudine (AZT) plus didanosine (ddI), AZT plus zalcitabine (ddC), AZT alternating monthly with ddI, and AZT/ddI plus nevirapine, work in AIDS patients whose CD4 counts were less than or equal to 50 cells/mm3. We are interested in evaluating the clinical efficacy of the four treatments using data of 1309 participants. After randomization, the CD4 counts of participants were scheduled to be collected at baseline and at 8-week intervals for a total length of 40 weeks. However, since participants had occasionally skipped or delayed visits and some dropped out halfway during the study, they had different numbers of measurements varied from one to nine with a median of four, which caused the study data to be unbalanced.

Methods: *Descriptive Analysis:* To ensure there is no significant difference in baseline characteristics in patients after randomizing them into the four treatment groups, we display the gender distribution, and the mean and standard deviation of the baseline age and log-CD4 count of four treatments groups in Table 1. Figure 1 will show the trend of log-CD4 counts over time for four treatment groups. Based on the design of the study, each participant should have 6 observations in 40-weeks, so in this analysis, we define participants who have less than six observations as those with missing data. Table 2 summarizes the baseline characteristics comparing participants with missing data to those who have complete observations.

Confirmatory Analysis: To characterize treatment effectiveness over time, we want to know if every treatment has the same rate of change in log-CD4 counts over time. Since this is an unbalanced data set, we decide to fit the data with a linear mixed model with random intercept and slope to allow different patients to follow their own line. Considering the majority of patients participated in follow-up approximately once every 8 weeks, which was too short for CD4 count to fluctuate drastically, we choose to use a correlated linear mixed model. Since the number of patients is much larger than the number of coefficients, use of REML and ML does not vary much, but we decided to use ML, which would allow us to use a Likelihood Ratio Test to test if the coefficients for all interaction terms equal to 0. Also, since previous studies indicated a change point of log-CD4 counts at week 16, and Figure 1 demonstrates that as well, we would use a linear spline model with a change point at week 16 for time. Thus, the final model is presented as below:

$$E[Y_{ij}|X_{ij}] = \beta_0 + \beta_1 t_{ij} + \beta_2 (t_{ij} - 16)^+ + \beta_3 I_{group=2} + \beta_4 I_{group=3} + \beta_5 I_{group=4} + \beta_6 I_{group=2} \times t_{ij} + \beta_7 I_{group=3} \times t_{ij} + \beta_8 I_{group=4} \times t_{ij} + \beta_9 I_{group=2} \times (t_{ij} - 16)^+ + \beta_{10} I_{group=3} \times (t_{ij} - 16)^+ + \beta_{11} I_{group=4} \times (t_{ij} - 16)^+$$

We also want to know if treatment effectiveness varies over time controlling for gender and baseline age, so we would use the same linear mixed model but with additional gender and baseline age covariates to answer this question.

Results: *Descriptive Analysis:* For each treatment group, a similar amount of about 327 participants was assigned. The mean and standard deviation of the age, number of visits, and baseline log-CD4 count are also similar in these four groups, except the AZT/ddI plus nevirapine group seems to have a slightly lower mean baseline log-CD4 count compared to the other three treatment groups. The proportion of male participants is also very similar in all four treatment groups (Table 1). From Table 2, we can see the characteristics between participants with missing data and participants with complete data are similar in baseline log-CD4 count and age. It seems that the group of participants who have missing data has a higher proportion of female participants. Also, the participants inside the AZT/ddI plus nevirapine group tend to have a higher proportion of participants with complete observations than other treatment groups, with AZT alternating monthly with ddI group having a lower proportion.

From Figure 1, we can see that the log-CD4 count tends to decrease over time for four groups generally. The log-CD4 count of AZT/ddI plus nevirapine group seems to increase from baseline until around week 16, and starts decreasing, when the log-CD4 counts of the other three groups are constantly decreasing over time. Also, the AZT/ddI plus nevirapine group seems to have the most minor declination in log-CD4 count between baseline and week 40, and the AZT alternating monthly with ddI group has the most significant declination. After week 8, the log-CD4 count is the highest in AZT/ddI plus nevirapine group, the second-highest in the AZT plus ddI group, and similar in the rest two groups. The evidence of positive correlation also can be found in Figure 1. The participants who have a high log-CD4 count in the first few weeks tend to stay high over the whole period, and vice versa. There are some exceptional cases in our data: the log-CD4 counts of certain participants suddenly drop to zero and increase back in the following measurement. It may be due to the measurement errors that accidentally measured the log-CD4 count to be zero. Except for these cases, the positive correlation obviously holds in Figure 1.

Confirmatory Analysis: There is statistically significant evidence at the 0.05 level that the rate of change in log-CD4 counts differs across different treatments before or after 16 weeks, either controlling for gender and baseline age or not (LRT $p < 0.001$). Based on this conclusion, we listed point estimates and 95% confidence intervals (CI) for the rate of change before and after 16 weeks of each treatment group, for both unadjusted and adjusted models (Table 3). Since this data did not provide a control group, we used zidovudine (AZT) alternating monthly with 400 mg didanosine (ddI) before week 16 as the reference group based on its overall lowest log-CD4 counts and the highest rate of declination in log-CD4 counts over time presented in Figure 1. From Table 3, in both of the models, the two groups that show statistically significant difference compared to AZT alternating with ddI before week 16 are AZT plus ddI before week 16, a 0.015 log-CD4 count increase (95% CI: 0.005, 0.024), and AZT/ddI plus nevirapine before week 16, a 0.035 log-CD4 count increase (95% CI: 0.026, 0.045). In the two models, we obtained very similar data, so we can conclude that, whether or not controlled for gender and baseline age, there is strong evidence that AZT plus ddI and AZT/ddI plus nevirapine before week 16 were relatively more effective than AZT alternating with ddI before week 16, but other treatments, both before and after 16 weeks, show no evidence of differences in effectiveness compared to AZT/ddI plus nevirapine before week 16.

Tables and Figures

Table 1: Baseline characteristics of patients by treatment group and overall

	AZT alternating ddI (N=325)	AZT + ddC (N=324)	AZT + ddI (N=330)	AZT/ddI + nevirapine (N=330)	Overall (N=1309)
Age (year)					
Mean (SD)	37.9 (8.59)	37.7 (7.90)	37.5 (8.06)	37.9 (8.26)	37.7 (8.20)
Sex					
Male	290 (89.2%)	282 (87.0%)	286 (86.7%)	289 (87.6%)	1147 (87.6%)
Female	35.0 (10.8%)	42.0 (13.0%)	44.0 (13.3%)	41.0 (12.4%)	162 (12.4%)
Baseline Log(CD4 count+1)					
Mean (SD)	2.98 (0.804)	2.93 (0.933)	2.91 (0.954)	2.84 (0.963)	2.91 (0.917)
Missing	5.00 (1.5%)	2.00 (0.6%)	3.00 (0.9%)	0 (0%)	10.0 (0.8%)
Number of Visits					
Mean (SD)	3.81 (1.51)	3.86 (1.54)	3.80 (1.55)	3.92 (1.52)	3.85 (1.53)

Figure 1: Loess Smooth line of log(CD4 count+1) overtime for four treatment groups

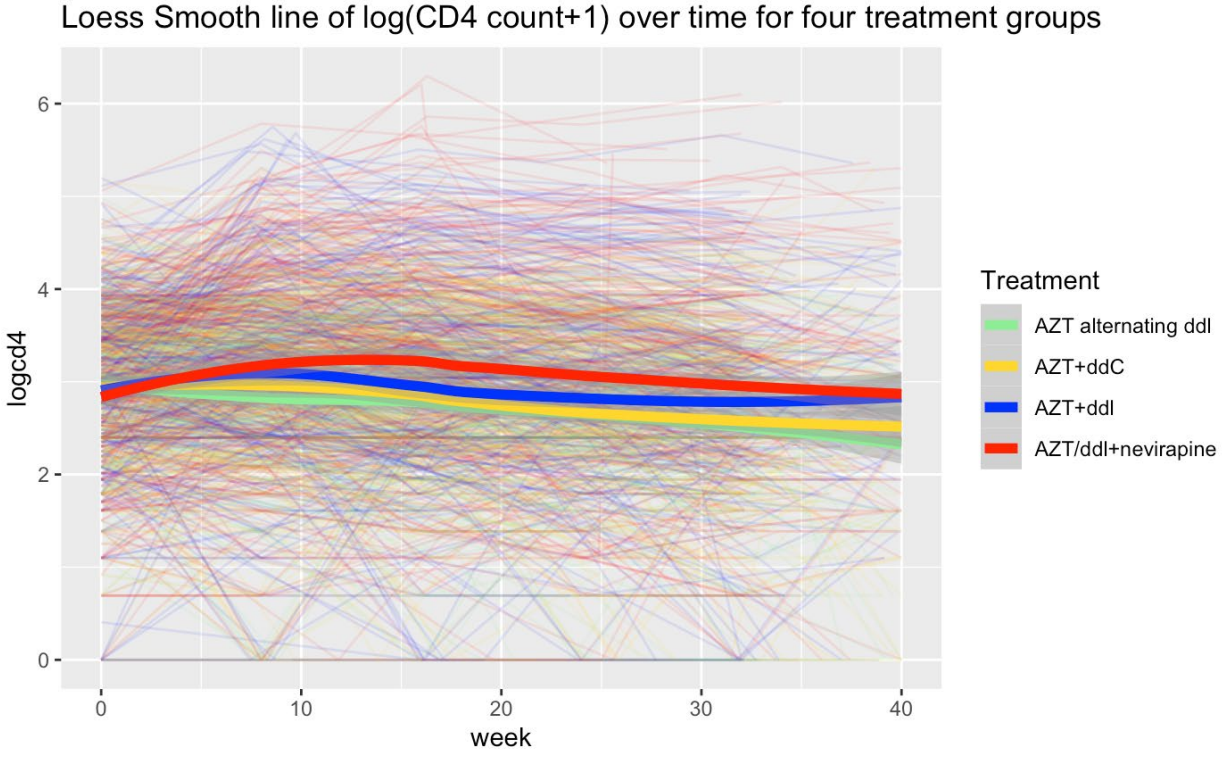

Table 2: Baseline demographic characteristics of patients by missing data status and overall

	Complete (N=183)	Missing (N=1126)	Overall (N=1309)
Age (year)			
Mean (SD)	38.0 (8.10)	37.7 (8.22)	37.7 (8.20)
Sex			
Male	168 (91.8%)	979 (86.9%)	1147 (87.6%)
Female	15.0 (8.2%)	147 (13.1%)	162 (12.4%)
Baseline Log(CD4 count+1)			
Mean (SD)	2.97 (0.846)	2.90 (0.928)	2.91 (0.917)
Missing	0 (0%)	10.0 (0.9%)	10.0 (0.8%)
Number of Visits			
Mean (SD)	6.04 (0.274)	3.49 (1.34)	3.85 (1.53)
Treatment Group			
AZT alternating ddI	40.0 (21.9%)	285 (25.3%)	325 (24.8%)
AZT+ddC	48.0 (26.2%)	276 (24.5%)	324 (24.8%)
AZT+ddI	42.0 (23.0%)	288 (25.6%)	330 (25.2%)
AZT/ddI+nevirapine	53.0 (29.0%)	277 (24.6%)	330 (25.2%)

Table 3: Point estimates (95% CI) for the rate of change before and after 16 weeks of each treatment group with AZT alternating ddI before week 16 as reference group for unadjusted and adjusted (for gender and baseline age) models

	Unadjusted	P-value	Adjusted for gender and baseline age	P-value
Before Week 16				
AZT alternating ddI	Reference		Reference	
AZT + ddC	0.0065 (-0.0032, 0.016)	0.19	0.0064 (-0.0032, 0.016)	0.19
AZT + ddI	0.015 (0.0053, 0.024)	0.0023	0.015 (0.0053, 0.024)	0.0023
AZT/ddI + nevirapine	0.035 (0.026, 0.045)	< 0.0001	0.035 (0.026, 0.045)	< 0.0001
After Week 16				
AZT alternating ddI	-0.0039 (-0.015, 0.0070)	0.48	-0.0039 (-0.015, 0.0070)	0.48
AZT + ddC	-0.0050 (-0.014, 0.0040)	0.28	-0.0050 (-0.014, 0.0040)	0.27
AZT + ddI	-0.0047 (-0.014,	0.31	-0.0047 (-0.014,	0.31

	0.0044)		0.0043)	
AZT/ddI + nevirapine	-0.0053 (-0.014, 0.0036)	0.24	-0.0054 (-0.014, 0.0035)	0.24

Code Appendix

```
library(uwIntroStats)
library(nlme)
library(dplyr)
library(reshape2)
library(ggplot2)
library(joineR)
library(MASS)
library(lattice)
#Read Data
cd4 <- read.csv("~/Desktop/R hw/cd4.csv")
cd4 = cd4[,-1]
cd4$treatment_group = as.factor(cd4$group)

#Data Organizing
unique_logcd4 = unique(cd4[,c("id","sex","age","group")])
byid = cd4 %>% group_by(id) %>% filter(week == 0)
byid = byid[,c(1,6)]
summary = by(unique_logcd4[,c(2,3)], INDICES = unique_logcd4$group, FUN = descrip)
visit_counts <- cd4 %>% group_by(id) %>% summarise(n=n(), nobs=sum(!is.na(logcd
4)))
table(visit_counts$nobs)
n_visit = cd4%>% group_by(id) %>% summarize(n())


cd4_demo = cd4 %>% group_by(id) %>% do(head(.,1))
cd4_demo = merge(cd4_demo, n_visit, all.x = T, by ="id")
cd4_demo = cd4_demo[,-6]
cd4_demo = merge(cd4_demo, byid, all.x = T, by= "id")
cd4_demo$missing = NA
cd4_demo$missing[cd4_demo$`n()`<6] = "Missing"
cd4_demo$missing[cd4_demo$`n()`>=6] = "Complete"
cd4_demo$sex <- factor(cd4_demo$sex, levels = c(1,0),
                       labels = c("Male", "Female"))
label(cd4_demo$age) <- "Age (year)"
label(cd4_demo$sex) <- "Sex"
label(cd4_demo$logcd4) <- "Baseline Log(CD4 count+1)"
label(cd4_demo$`n()`) <- "Number of Visits"
label(cd4_demo$treatment_group) = "Treatment Group"
cd4_demo$treatment_group = factor(cd4_demo$group, levels = c(1,2,3,4),
                                  labels = c("AZT alternating ddl",
                                             "AZT+ddC","AZT+ddl","AZT/ddl+nevirapi
ne"))
```

```r
label(cd4_demo$treatment_group) = "Treatment Group"

#Figure 1
ggplot(cd4, aes(x=week, y=logcd4, group=id, color=as.factor(group))) +
  geom_line(aes(color=as.factor(group)), alpha=0.1) +
  geom_smooth(method = 'loess', aes(group=NULL,color=as.factor(group)), size=2)+
  scale_color_manual(values=c("light green", "gold", "blue","red"),
                                          labels = c("AZT alternating ddl",
                                                     "AZT+ddC","AZT+ddl","AZT/ddl+n
evirapine"),
                                          name = "Treatment")+
  labs(title = "Loess Smooth line of log(CD4 count+1) over time for four treatment
 groups")



#Table 1
table1(~ age + sex + logcd4 + `n()`| group, data=cd4_demo,
       render.continuous=c(.="Mean (SD)"))

#Table 2
table1(~ age + sex + logcd4 + `n()` + treatment_group| missing, data=cd4_demo,
       render.continuous=c(.="Mean (SD)"))



### Models
# Random intercept + slope + spline at week 16
cd4$weekSpline16 <- (cd4$week-16)*(cd4$week>16)
cd4$group <- factor(cd4$group, levels = c(1,2,3,4),
                          labels = c("I","II", "III", "IV"))

mod <- lme(fixed=logcd4 ~ (week + weekSpline16)*group,
            method="ML", data=cd4, na.action=na.omit,
            random=reStruct(~1 + week + weekSpline16 | id,REML=F))
summary(mod)
intervals(mod)

sum.dat <- as.data.frame(coef(summary(mod)))

mod_reduced <- lme(fixed=logcd4 ~ week + weekSpline16+group, method="ML",
                   random=reStruct(~1 + week + weekSpline16 | id, REML=F), data=c
d4,
                   na.action=na.omit)
anova(mod, mod_reduced)

mod_1_16 <- glht(mod, linfct=c("weekSpline16 = 0"))
confint(mod_1_16)
summary(mod_1_16)

mod_2_16 <- glht(mod, linfct=c("weekSpline16 + week:groupII + weekSpline16:groupII
 = 0"))
```

```r
summary(mod_2_16)
confint(mod_2_16)

mod_3_16 <- glht(mod, linfct=c("weekSpline16 + week:groupIII + weekSpline16:groupI
II = 0"))
summary(mod_3_16)
confint(mod_3_16)

mod_4_16 <- glht(mod, linfct=c("weekSpline16 + week:groupIV + weekSpline16:groupIV
 = 0"))
summary(mod_4_16)
confint(mod_4_16)

# Control for age and gender
cd4$sex <- factor(cd4$sex, levels = c(1,0),
                  labels = c("Male","Female"))

mod_c <- lme(fixed=logcd4 ~ (week + weekSpline16)*group + sex+age,
           method="ML", data=cd4, na.action=na.omit,
           random=reStruct(~1 + week + weekSpline16 | id,REML=F))
summary(mod_c)
intervals(mod_c)

sum.c.dat <- as.data.frame(coef(summary(mod_c)))

mod_c_reduced <- lme(fixed=logcd4 ~ week + weekSpline16+group+ sex+age, method="ML
",
                   random=reStruct(~1 + week + weekSpline16 | id, REML=F), data=cd
4,
                   na.action=na.omit)
anova(mod_c, mod_c_reduced)

mod_2_16 <- glht(mod_c, linfct=c("weekSpline16 + week:groupII + weekSpline16:group
II = 0"))
summary(mod_2_16)
confint(mod_2_16)

mod_3_16 <- glht(mod_c, linfct=c("weekSpline16 + week:groupIII + weekSpline16:grou
pIII = 0"))
summary(mod_3_16)
confint(mod_3_16)

mod_4_16 <- glht(mod_c, linfct=c("weekSpline16 + week:groupIV + weekSpline16:group
IV = 0"))
summary(mod_4_16)
confint(mod_4_16)
```