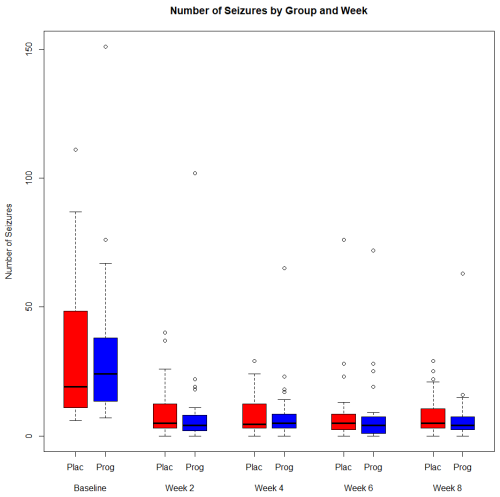
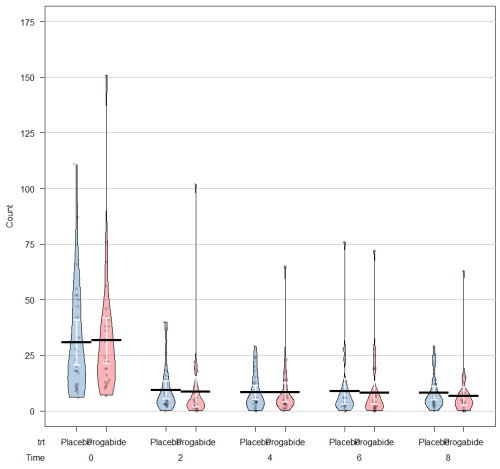
GEE and GLMM

2022-04-18

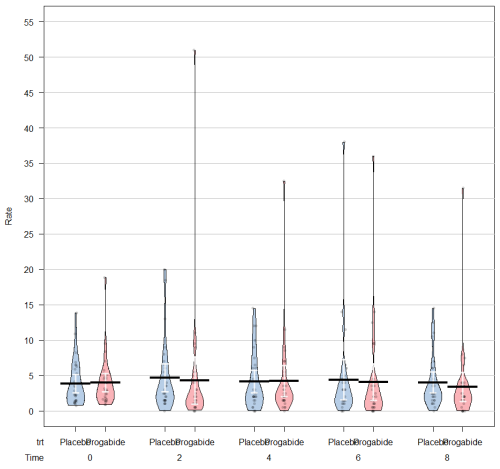
```
##
### GLMMs and GEEs with Epilepsy Data
##

##### Load Required Libraries #####
library(tidyverse)  # ggplot2, tidyr
library(lme4)     # Functions: glmer
library(gee) # Functions: gee

##
#### Read in data set into R:
##
epilepsy <-
read.table("http://www.hsph.harvard.edu/fitzmaur/ala2e/epilepsy.dat",
                        header=FALSE)
names(epilepsy) <-
c("ID","trt","age","Week0","Week2","Week4","Week6","Week8")
epilepsy$trt <- factor(epilepsy$trt, levels=c(0,1),
labels=c("Placebo","Progabide"))

## Convert to Long form:
epi_long <- pivot_longer(epilepsy,
                         cols = 4:8,
                         names_to = "Time", names_prefix = "Week",
                         values_to = "Count")
epi_long$Time <- as.numeric(epi_long$Time)

## Create new variables
epi_long <- epi_long %>% mutate(
  PostBase = as.numeric(Time != 0),
  Weeks = 8*(PostBase==0) + 2*(PostBase==1)
)


##### Exploratory Data Analysis #####

## Number of seizures by group by time period:
boxplot(Count ~ trt+Time, at=c(1,2,4,5,7,8,10,11,13,14), data=epi_long,
col=c("red","blue"), ylab="Number of Seizures", xlab="",
names=rep(c("Plac","Prog"),5), main="Number of Seizures by Group and Week")
mtext( c("Baseline",paste("Week",c(2,4,6,8))), side=1,
at=c(1.5,4.5,7.5,10.5,13.5), line=3 )
```



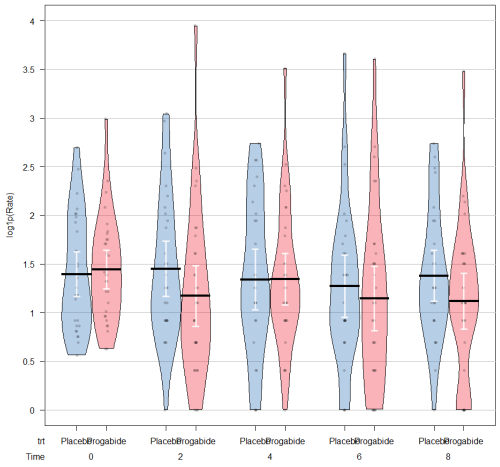

```
# Calculate rates per week:
epi_long <- epi_long %>% mutate(
  Rate = case_when(
    Time == 0 ~ Count/8,
    Time != 0 ~ Count/2
  )
)

library(yarrr)
pirateplot(Count ~ trt+Time, data = epi_long, inf.method = "ci", inf.disp =
"line")
```



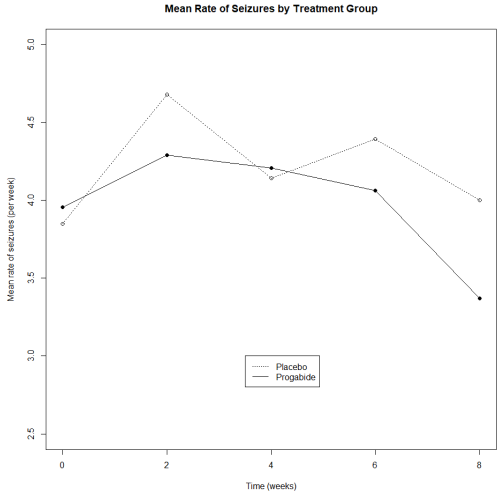

```
pirateplot(Rate ~ trt+Time, data = epi_long, inf.method = "ci", inf.disp =
"line")
```



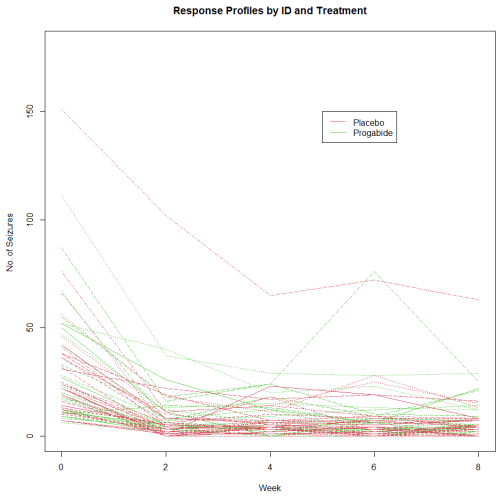

```
pirateplot(log1p(Rate) ~ trt+Time, data = epi_long, inf.method = "ci",
inf.disp = "line")
```



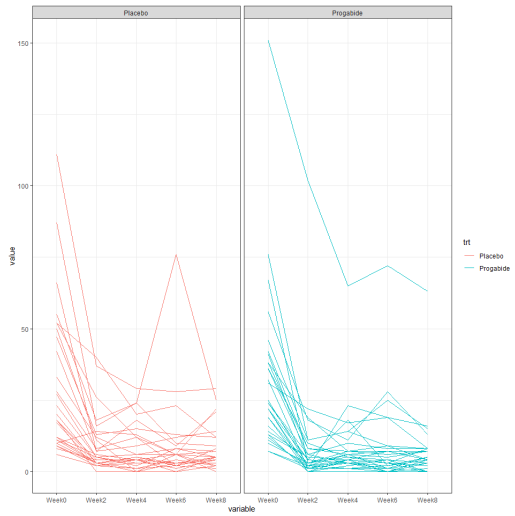

```
# Plot of mean rates:
means <- tapply(epi_long$Rate, list(epi_long$Time,epi_long$trt), mean)
matplot(matrix(c(0,2,4,6,8)), means,
        col=c(1,1), lty=c(3,1), type="o",
        pch=c(1,16), xlab="Time (weeks)",
        ylab="Mean rate of seizures (per week)",
        ylim=c(2.5,5.0),
        main="Mean Rate of Seizures by Treatment Group")
legend(3.5, 3.0, c("Placebo","Progabide"), lty=c(3,1))
```



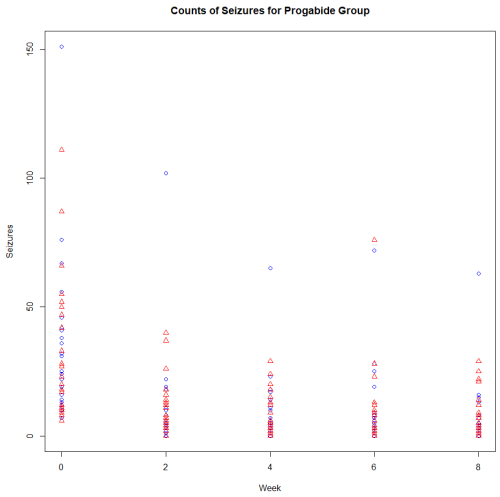

```
## Plots of individual counts:
matplot(matrix(c(0,2,4,6,8)), t(epilepsy[,4:8]),
    col=(as.numeric(epilepsy$trt=="Placebo")+2), type="l", ylim = c(0,180),
    xlab="Week",ylab="No. of Seizures ", main="Response Profiles by ID and
Treatment")
legend(5,150,c("Placebo","Progabide"),col=c(2,3),lty=c(1,1))
```




```
library(GGally)

epilepsy %>% ggparcoord(4:8, scale = "globalminmax", group = "trt") +
  theme_bw() +
  facet_wrap(~trt)
```




```
# Who is the outlier at baseline in the Progabide group?
plot(Count[trt=="Progabide"] ~ Time[trt=="Progabide"],
     xlab="Week", ylab="Seizures",
     main="Counts of Seizures for Progabide Group", col="blue",
     data=epi_long)
points(Count[trt=="Placebo"] ~ Time[trt=="Placebo"], pch=2,
       col="red", data=epi_long)
identify(epi_long$Count ~ epi_long$Time, labels=epi_long$ID)
```




```
## integer(0)
```

```
# ID 49 data:
epi_long[epi_long$ID==49,]
```

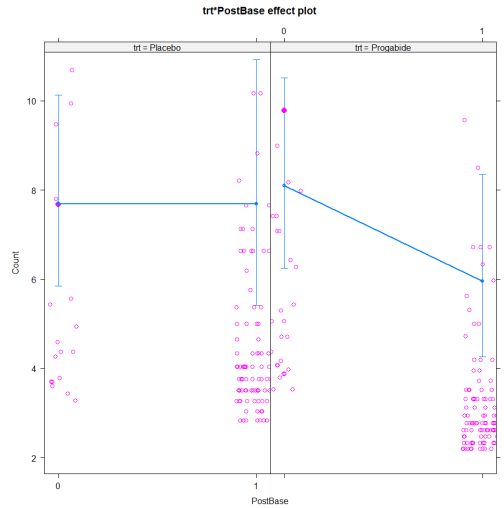
```
## # A tibble: 5 x 8
##      ID trt         age  Time Count PostBase Weeks  Rate
##   <int> <fct>     <int> <dbl> <int>    <dbl> <dbl> <dbl>
## 1    49 Progabide    22     0   151        0     8  18.9
## 2    49 Progabide    22     2   102        1     2  51
## 3    49 Progabide    22     4    65        1     2  32.5
## 4    49 Progabide    22     6    72        1     2  36
## 5    49 Progabide    22     8    63        1     2  31.5
```

```
# This patient could have a large impact on the analysis -
# Book does analysis with and without ID 49.
```

```
##### Fitting GLMMs #####
# Since the number of weeks each count refers to differs
# (8 weeks for baseline by 2 weeks afterwards)
# we need to include an "offset" --> Model mean rate per week

# With glmer (and lmer) function, random effects specified in parentheses:
?lmer
?glmer

## Fit GLMM
epi_long <- epi_long %>% mutate(PostBase = factor(PostBase))

mod1 <- glmer(Count ~ trt*PostBase + (PostBase | ID), offset=log(Weeks),
              family=poisson, data=epi_long)
summary(mod1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: Count ~ trt * PostBase + (PostBase | ID)
##    Data: epi_long
##  Offset: log(Weeks)
##
##      AIC      BIC   logLik deviance df.resid
##   1864.4   1890.2   -925.2   1850.4      288
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1394 -0.7073 -0.0620  0.5138  6.9653
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.4999   0.7070
##         PostBase1   0.2319   0.4816   0.16
## Number of obs: 295, groups:  ID, 59
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                1.0708453  0.1402715   7.634 2.27e-14
## trtProgabide               0.0512167  0.1927137   0.266   0.7904
## PostBase1                 -0.0004996  0.1091005  -0.005   0.9963
## trtProgabide:PostBase1 -0.3062158  0.1504204  -2.036   0.0418
##
## Correlation of Fixed Effects:
##             (Intr) trtPrg PstBs1
## trtProgabid -0.725
## PostBase1    0.011 -0.013
## trtPrgb:PB1 -0.014  0.025 -0.709
```

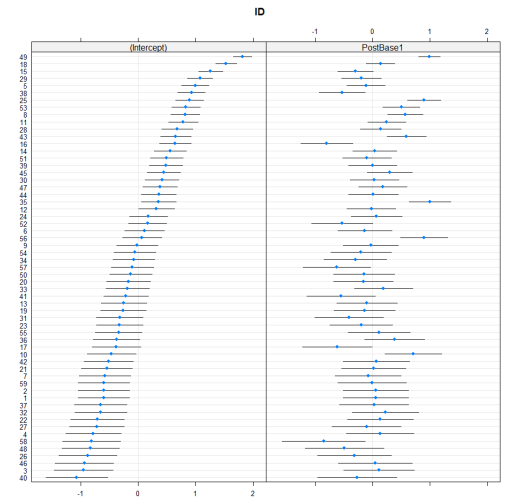
```
plot(allEffects(mod1, residuals = T), type = "response", x.var = "PostBase")
#Issues in scaling with offset in Poisson rate models
```



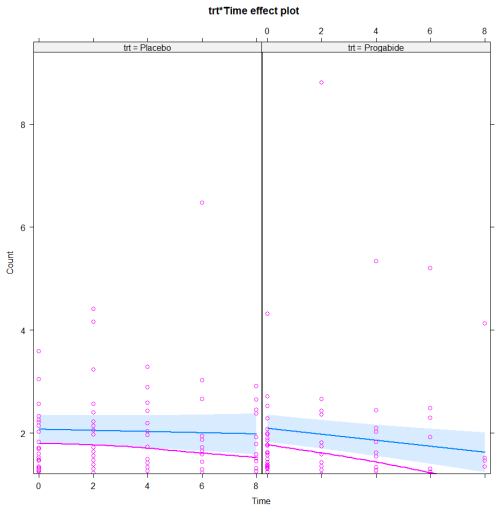

```
# Estimated random effects
ranef(mod1)
```

```
## $ID
##    (Intercept)    PostBase1
## 1  -0.60197368  0.053075961
## 2  -0.60197368  0.053075961
## 3  -0.95584894  0.112780219
## 4  -0.78751956  0.127425364
## 5   0.99154663 -0.116149180
## 6   0.10948363 -0.138443641
## 7  -0.57934911 -0.079468569
## 8   0.81464415  0.566916764
## 9  -0.01965844 -0.032916035
## 10 -0.46624571  0.706002887
## 11  0.78554645  0.243655497
## 12  0.31770399 -0.025038532
## 13 -0.25263814 -0.103699520
## 14  0.55723002  0.033499564
## 15  1.26048140 -0.300455317
## 16  0.64351430 -0.804224630
## 17 -0.38232599 -0.620312869
## 18  1.52980608  0.133760131
## 19 -0.26255896 -0.145316442
## 20 -0.17314632 -0.164548928
## 21 -0.54995950  0.015238661
## 22 -0.71194461  0.132566542
## 23 -0.32687377 -0.202650257
## 24  0.17754869  0.066827810
## 25  0.89477469  0.893345733
## 26 -0.88246585 -0.321624005
## 27 -0.72517425 -0.108837284
## 28  0.67746018  0.137684042
## 29  1.07892366 -0.202875872
## 30  0.41195329  0.028095355
## 31 -0.32260666 -0.412393859
## 32 -0.65611911  0.221362649
## 33 -0.18652492  0.186766976
## 34 -0.07899798 -0.305198736
## 35  0.35362987  0.996654913
## 36 -0.37841127  0.381263989
## 37 -0.65496455  0.021082520
## 38  0.92628754 -0.532610674
## 39  0.48171434 -0.001086374
## 40 -1.07529126 -0.273473822
## 41 -0.21504353 -0.556473959
## 42 -0.51869849  0.063582691
## 43  0.65368382  0.589958962
## 44  0.35793215  0.010241644
## 45  0.44457602  0.295731199
## 46 -0.94074737  0.042165812
## 47  0.37897839  0.173507520
## 48 -0.83298727 -0.495214573
## 49  1.81733174  0.986679577
## 50 -0.12992207 -0.153557315
## 51  0.49339328 -0.103671457
## 52  0.16410223 -0.536852808
## 53  0.82987828  0.497735332
## 54 -0.05966416 -0.204850414
## 55 -0.34384372  0.112038436
## 56  0.06474997  0.892379194
## 57 -0.10209408 -0.626688189
## 58 -0.81580027 -0.856342299
## 59 -0.60188744 -0.013537361
```

```
##
## with conditional variances for "ID"

library(lattice)
dotplot(ranef(mod1, postVar = T))

## $ID
```




```
# Subject-specific coefficients (beta_k + b_ki)
coef(mod1)

## $ID
##     (Intercept) trtProgabide    PostBase1 trtProgabide:PostBase1
## 1   0.468871645    0.0512167  0.052576367             -0.3062158
## 2   0.468871645    0.0512167  0.052576367             -0.3062158
## 3   0.114996378    0.0512167  0.112280625             -0.3062158
## 4   0.283325764    0.0512167  0.126925771             -0.3062158
## 5   2.062391948    0.0512167 -0.116648773             -0.3062158
## 6   1.180328954    0.0512167 -0.138943234             -0.3062158
## 7   0.491496206    0.0512167 -0.079968162             -0.3062158
## 8   1.885489471    0.0512167  0.566417170             -0.3062158
## 9   1.051186885    0.0512167 -0.033415628             -0.3062158
## 10  0.604599609    0.0512167  0.705503293             -0.3062158
## 11  1.856391768    0.0512167  0.243155903             -0.3062158
## 12  1.388549312    0.0512167 -0.025538125             -0.3062158
## 13  0.818207178    0.0512167 -0.104199113             -0.3062158
## 14  1.628075336    0.0512167  0.032999971             -0.3062158
## 15  2.331326722    0.0512167 -0.300954911             -0.3062158
## 16  1.714359623    0.0512167 -0.804724224             -0.3062158
## 17  0.688519329    0.0512167 -0.620812462             -0.3062158
## 18  2.600651396    0.0512167  0.133260537             -0.3062158
## 19  0.808286356    0.0512167 -0.145816035             -0.3062158
## 20  0.897699000    0.0512167 -0.165048522             -0.3062158
## 21  0.520885822    0.0512167  0.014739067             -0.3062158
## 22  0.358900710    0.0512167  0.132066949             -0.3062158
## 23  0.743971552    0.0512167 -0.203149851             -0.3062158
## 24  1.248394014    0.0512167  0.066328217             -0.3062158
## 25  1.965620009    0.0512167  0.892846140             -0.3062158
## 26  0.188379471    0.0512167 -0.322123598             -0.3062158
## 27  0.345671067    0.0512167 -0.109336877             -0.3062158
## 28  1.748305497    0.0512167  0.137184449             -0.3062158
## 29  2.149768985    0.0512167 -0.203375465             -0.3062158
## 30  1.482798610    0.0512167  0.027595762             -0.3062158
## 31  0.748238665    0.0512167 -0.412893453             -0.3062158
## 32  0.414726214    0.0512167  0.220863055             -0.3062158
## 33  0.884320404    0.0512167  0.186267382             -0.3062158
## 34  0.991847337    0.0512167 -0.305698330             -0.3062158
## 35  1.424475185    0.0512167  0.996155320             -0.3062158
## 36  0.692434055    0.0512167  0.380764395             -0.3062158
## 37  0.415880775    0.0512167  0.020582926             -0.3062158
## 38  1.997132857    0.0512167 -0.533110268             -0.3062158
## 39  1.552559660    0.0512167 -0.001585967             -0.3062158
## 40 -0.004445941    0.0512167 -0.273973415             -0.3062158
## 41  0.855801793    0.0512167 -0.556973553             -0.3062158
## 42  0.552146835    0.0512167  0.063083098             -0.3062158
## 43  1.724529145    0.0512167  0.589459369             -0.3062158
## 44  1.428777467    0.0512167  0.009742051             -0.3062158
## 45  1.515421335    0.0512167  0.295231605             -0.3062158
## 46  0.130097946    0.0512167  0.041666218             -0.3062158
## 47  1.449823713    0.0512167  0.173007927             -0.3062158
## 48  0.237858052    0.0512167 -0.495714167             -0.3062158
## 49  2.888177061    0.0512167  0.986179984             -0.3062158
## 50  0.940923250    0.0512167 -0.154056908             -0.3062158
## 51  1.564238598    0.0512167 -0.104171051             -0.3062158
## 52  1.234947552    0.0512167 -0.537352401             -0.3062158
## 53  1.900723598    0.0512167  0.497235739             -0.3062158
## 54  1.011181164    0.0512167 -0.205350008             -0.3062158
## 55  0.727001602    0.0512167  0.111538842             -0.3062158
## 56  1.135595294    0.0512167  0.891879600             -0.3062158
## 57  0.968751238    0.0512167 -0.627187783             -0.3062158
## 58  0.255045045    0.0512167 -0.856841892             -0.3062158
## 59  0.468957884    0.0512167 -0.014036954             -0.3062158
##
## attr(,"class")
## [1] "coef.mer"

# Treating time as quantitative:
mod2 <- glmer(Count ~ trt*Time + (Time | ID), offset=log(Weeks),
              family=poisson, data=epi_long)

summary(mod2)

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: Count ~ trt * Time + (Time | ID)
##    Data: epi_long
##  Offset: log(Weeks)
##
##      AIC      BIC   logLik deviance df.resid
##   1924.2   1950.0   -955.1   1910.2      288
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3786 -0.7228 -0.1173  0.5846  6.6309
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.526863 0.72585
##         Time        0.005029 0.07091  0.22
## Number of obs: 295, groups:  ID, 59
##
## Fixed effects:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.10395    0.14260   7.741 9.83e-15
## trtProgabide      0.01750    0.19632   0.089   0.9290
## Time             -0.01133    0.01681  -0.674   0.5004
## trtProgabide:Time -0.04669    0.02335  -2.000   0.0456
##
## Correlation of Fixed Effects:
##             (Intr) trtPrg Time
## trtProgabid -0.724
## Time         0.065 -0.053
## trtPrgbd:Tm -0.054  0.074 -0.694

plot(allEffects(mod2, residuals = T), type = "link")
```




```
# How does the interpretation of coefficients change?
# Note larger AIC


##### Fitting Marginal Models #####
?gee
mod.gee <- gee(Count ~ trt*PostBase + offset(log(Weeks)),
               id = ID, family = poisson(link = "log"),
               corstr = "exchangeable", data = epi_long)

##           (Intercept)          trtProgabide              PostBase1
##            1.34760922            0.02753449            0.11183602
## trtProgabide:PostBase1
##           -0.10472579
```

```
# corstr="exchangeable" --> compound symmetry covariance structure.
# family=poisson --> Poisson variance function (not distribution)
summary(mod.gee)

##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                      Logarithm
##  Variance to Mean Relation: Poisson
##  Correlation Structure:     Exchangeable
##
## Call:
## gee(formula = Count ~ trt * PostBase + offset(log(Weeks)), id = ID,
##     data = epi_long, family = poisson(link = "log"), corstr =
"exchangeable")
##
## Summary of Residuals:
##        Min         1Q     Median         3Q        Max
##  -4.303571  -1.303571   2.016129  10.370392 147.044355
##
##
## Coefficients:
##                            Estimate Naive S.E.    Naive z Robust S.E.
Robust z
## (Intercept)             1.34760922  0.1510969  8.9188397   0.1573571
8.5640166
## trtProgabide            0.02753449  0.2071018  0.1329515   0.2217878
0.1241479
## PostBase1               0.11183602  0.1545145  0.7237900   0.1159304
0.9646821
## trtProgabide:PostBase1 -0.10472579  0.2197052 -0.4766650   0.2134448 -
0.4906459
##
## Estimated Scale Parameter:  19.6797
## Number of Iterations:  1
##
## Working Correlation
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 1.0000000 0.7713861 0.7713861 0.7713861 0.7713861
## [2,] 0.7713861 1.0000000 0.7713861 0.7713861 0.7713861
## [3,] 0.7713861 0.7713861 1.0000000 0.7713861 0.7713861
## [4,] 0.7713861 0.7713861 0.7713861 1.0000000 0.7713861
## [5,] 0.7713861 0.7713861 0.7713861 0.7713861 1.0000000

mod.gee2 <- gee(Count ~ trt*PostBase + offset(log(Weeks)),
                id = ID, family = poisson(link = "log"),
                corstr = "AR-M", data = epi_long)

##           (Intercept)           trtProgabide               PostBase1
##            1.34760922              0.02753449              0.11183602
## trtProgabide:PostBase1
##            -0.10472579

summary(mod.gee2)

##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                      Logarithm
##  Variance to Mean Relation: Poisson
##  Correlation Structure:     AR-M , M = 1
##
## Call:
## gee(formula = Count ~ trt * PostBase + offset(log(Weeks)), id = ID,
##     data = epi_long, family = poisson(link = "log"), corstr = "AR-M")
##
## Summary of Residuals:
##        Min         1Q     Median         3Q        Max
##  -4.327892  -1.327892   2.120474  10.440487 147.208867
##
##
## Coefficients:
##                            Estimate Naive S.E.    Naive z Robust S.E.
## (Intercept)             1.31279985  0.1427491  9.1965551   0.1617122
## trtProgabide            0.01986517  0.1960117  0.1013468   0.2117125
## PostBase1               0.15228086  0.1682744  0.9049558   0.1114624
## trtProgabide:PostBase1 -0.12923296  0.2405021 -0.5373464   0.2597892
##                            Robust z
## (Intercept)             8.11812310
## trtProgabide            0.09383086
## PostBase1               1.36620870
## trtProgabide:PostBase1 -0.49745325
##
## Estimated Scale Parameter:  20.12528
## Number of Iterations:  3
##
## Working Correlation
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 1.0000000 0.8102249 0.6564644 0.5318838 0.4309455
## [2,] 0.8102249 1.0000000 0.8102249 0.6564644 0.5318838
## [3,] 0.6564644 0.8102249 1.0000000 0.8102249 0.6564644
## [4,] 0.5318838 0.6564644 0.8102249 1.0000000 0.8102249
## [5,] 0.4309455 0.5318838 0.6564644 0.8102249 1.0000000

# Note similar coefficient estimates and Wald tests
# GEE std. errs robust to covariance structure assumption
```