# Code Appendix

Laura Cosgrove
12/4/2019

#### Contents

```
Read data
                                                                      1
Explore outcome distribution
                                                                      1
Exploratory Mean and Trajectory Plots
                                                                      3
GEE models
                                                                      8
  8
Fit GEE models
                                                                      9
library(tidyverse)
## -- Attaching packages -----
                 v purrr
## v ggplot2 3.2.1
                          0.3.2
               v dplyr
## v tibble 2.1.3
                         0.8.3
## v tidyr 1.0.0 v stringr 1.4.0
## v readr
        1.3.1
                 v forcats 0.4.0
## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()
                masks stats::lag()
library(geepack)
```

#### Read data

```
alc_dat <- readxl::read_xls("./data/ALCDEP.xls")

alc_dat <- janitor::clean_names(alc_dat) %>%
    mutate(treatment = factor(treatment)) %>%
    mutate(gender = factor(gender, levels = c(0,1), labels = c("Male", "Female")))

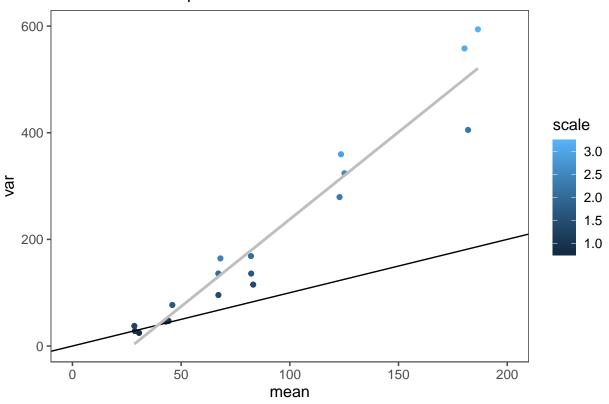
alc_dat_long <- alc_dat %>%
    pivot_longer(nd0:nd60, names_to = "days", values_to = "drinks", names_prefix = "nd")
```

### Explore outcome distribution

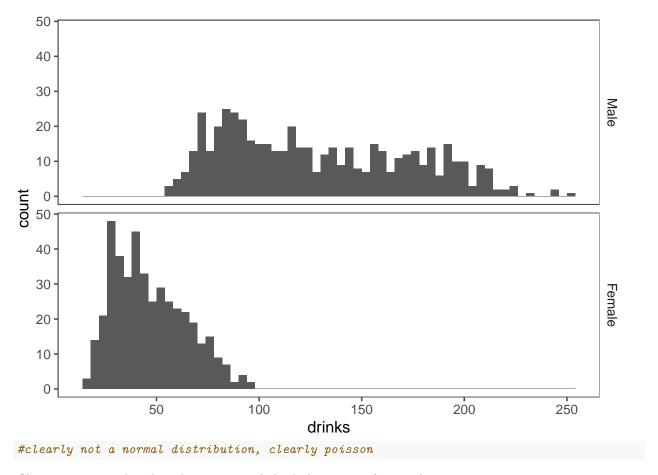
```
summary(alc_dat_long$drinks)

## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 16.00 47.00 79.00 92.01 127.00 252.00
```

# Potential overdispersion in Poisson outcome



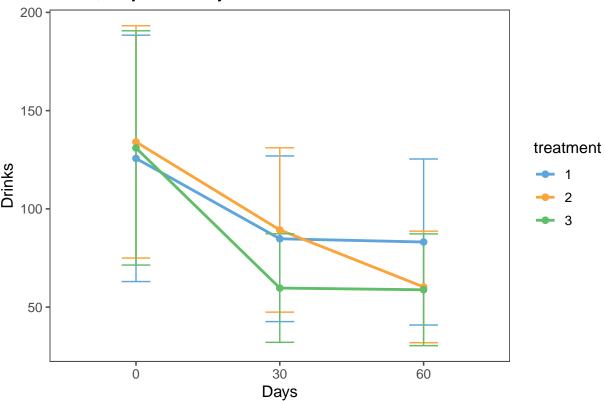
#still some overdispersion if poisson in the gender x treatment x time model; maybe need negative binom
#normal approximation?
alc\_dat\_long %>%
ggplot(aes(x = drinks)) + geom\_histogram(bins = 60) + ggthemes::theme\_few() + facet\_grid(c("gender"))



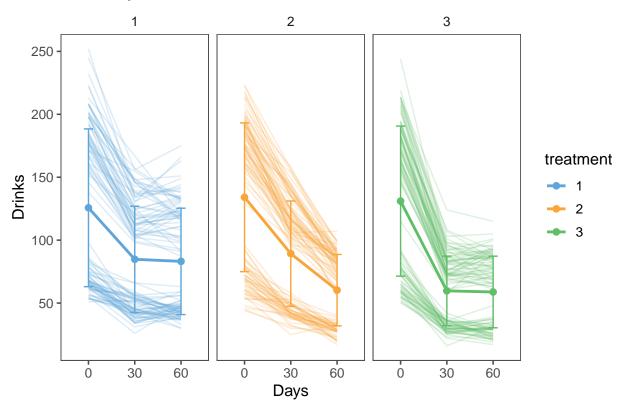
Choose a poisson-distributed outcome and check diagnostics for overdispersion.

# **Exploratory Mean and Trajectory Plots**

### Mean, Trajectories by Treatment



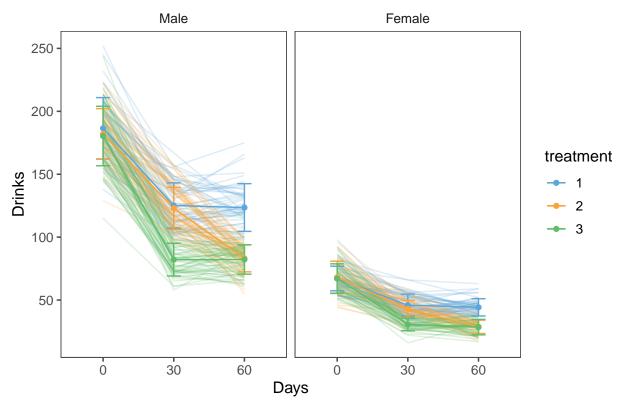
# **Raw Trajectories**



There are clearly two groups of patients in all groups, with a wide berth between the groups: those whose baseline number of alcoholic drinks in the past 30 days is less than 100 (~3 drinks/day) and those who are closer to 150 in the past 30 days (5 drinks a day).

Perhaps this could be a gender effect:

# **Raw Trajectories**



It is, in fact, a gender effect, and variance shrinks a great deal when the outcome is stratified on gender. Luckily, there does not seem to be heterogeneity of treatment effect by gender; the treatment trajectories look the same, but scaled to the respective within-gender mean.

Now, lets's look at faceting on relapsers vs. no relapsers:

```
library(janitor)

##

## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':

##

## chisq.test, fisher.test

alc_dat %>%

  tabyl(treatment, relapse) %>%
  adorn_percentages("row") %>%
  adorn_pert_formatting(digits = 2) %>%
  adorn_ns() %>%
  knitr::kable(caption = "Contingency table of Relapse by Treatment")
```

Table 1: Contingency table of Relapse by Treatment

treatment	0	1
1	17.92% (19)	82.08% (87)
2	54.21% (58)	45.79% (49)
3	86.14% (87)	13.86% (14)

```
alc_dat %>%
  tabyl(treatment, relapse) %>%
  chisq.test()
## Pearson's Chi-squared test
##
## data:
## X-squared = 96.71, df = 2, p-value < 2.2e-16
#significant differences. pairwise:
alc_dat %>%
 filter(treatment %in% 1:2) %>%
 mutate(treatment = forcats::fct_drop(treatment)) %>%
 tabyl(treatment, relapse) %>%
 chisq.test(.)
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: .
## X-squared = 28.815, df = 1, p-value = 7.962e-08
alc dat %>%
 filter(treatment %in% c(1,3)) %>%
 mutate(treatment = forcats::fct_drop(treatment)) %>%
 tabyl(treatment, relapse) %>%
 chisq.test(.)
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:
## X-squared = 93.61, df = 1, p-value < 2.2e-16
alc dat %>%
 filter(treatment %in% 2:3) %>%
 mutate(treatment = forcats::fct_drop(treatment)) %>%
 tabyl(treatment, relapse) %>%
 chisq.test(.)
##
## Pearson's Chi-squared test with Yates' continuity correction
## data:
## X-squared = 23.603, df = 1, p-value = 1.184e-06
#all pairwise differences are highly significant
alc_dat_grouped <- alc_dat %>%
 mutate(treatment = forcats::fct_relevel(treatment, "3")) %>%
  group_by(treatment, gender) %>%
 count(relapse) %>%
  pivot_wider(names_from = relapse, values_from = n, names_prefix = "relapse_")
```

```
response_table = cbind(alc_dat_grouped$relapse_1, alc_dat_grouped$relapse_0)
treatment <- alc_dat_grouped$treatment
glm_relapse_ref <- glm(response_table ~ treatment , family = binomial)

broom::tidy(glm_relapse_ref, exponentiate = TRUE, conf.int = TRUE) %>%
    dplyr::select(term, estimate, std.error, conf.low, conf.high) %>%
knitr::kable(caption = "Logistic Regression of Treatment on 6-month Relapse")
```

Table 2: Logistic Regression of Treatment on 6-month Relapse

term	estimate	std.error	conf.low	conf.high
(Intercept)	0.1609195	0.2879632	0.0876527	0.2734223
treatment1	28.4548872	0.3834692	13.8353111	62.5513004
treatment2	5.2500000	0.3472353	2.7182467	10.6829755

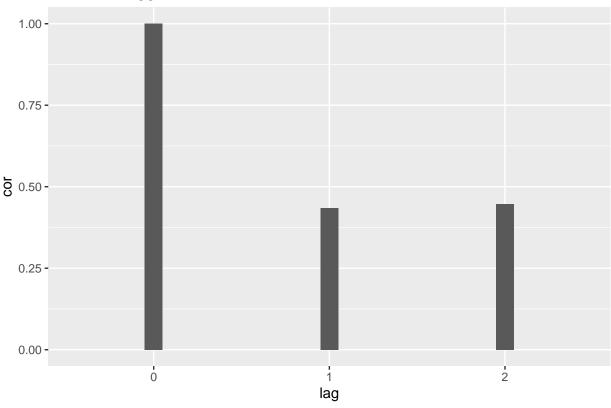
```
broom::glance(glm_relapse_ref) %>%
  summarize(test = 1 - pchisq(deviance, df.residual)) #acceptable
## # A tibble: 1 x 1
##
      test
##
     <dbl>
## 1 0.863
anova(glm_relapse_ref, test = "Chisq") #significant
## Analysis of Deviance Table
## Model: binomial, link: logit
##
## Response: response_table
## Terms added sequentially (first to last)
##
##
            Df Deviance Resid. Df Resid. Dev Pr(>Chi)
##
## NULL
                                5
                                     106.856
                                       0.745 < 2.2e-16 ***
## treatment 2
                 106.11
                                3
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
```

#### GEE models

#### **Exploratory Correlation Plots**

```
sat_mod <- glm(drinks ~ gender + treatment*days, family = "poisson", data = alc_dat_long) #super satura
alc_dat_long %>%
  modelr::add_residuals(sat_mod) %>%
  dplyr::select(sid, days, treatment, gender, resid) %>%
  mutate(days = as.numeric(days)) %>%
  group_by(sid) %>%
  mutate(lag_1 = lag(resid, 1),
```

#### ACF with lagged measurements



#### Fit GEE models

## geeglm(formula = drinks ~ treatment \* days + gender, family = poisson,

```
##
       data = alc_dat_long, id = sid, corstr = "exchangeable")
##
##
   Coefficients:
##
                     Estimate Std.err
                                            Wald Pr(>|W|)
## (Intercept)
                     5.22505 0.01566 1.114e+05
                                                   <2e-16 ***
                     -0.01568 0.01905 6.780e-01
                                                    0.410
## treatment2
## treatment3
                     -0.02564 0.02039 1.581e+00
                                                    0.209
## days30
                     -0.39386 0.01425 7.642e+02
                                                   <2e-16 ***
## days60
                     -0.41353 0.01253 1.089e+03
                                                   <2e-16 ***
## genderFemale
                    -1.00888 0.01448 4.856e+03
                                                   <2e-16 ***
## treatment2:days30 -0.01293 0.01901 4.620e-01
                                                    0.497
## treatment3:days30 -0.39227
                                                   <2e-16 ***
                              0.02087 3.534e+02
## treatment2:days60 -0.38624 0.01892 4.165e+02
                                                   <2e-16 ***
## treatment3:days60 -0.38748 0.01928 4.038e+02
                                                   <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## Estimated Scale Parameters:
              Estimate Std.err
##
## (Intercept)
                 1.901 0.1121
##
## Correlation: Structure = exchangeable Link = identity
##
## Estimated Correlation Parameters:
##
        Estimate Std.err
## alpha
          0.4603 0.03397
## Number of clusters:
                         314
                              Maximum cluster size: 3
pois_gee_int <- geeglm(drinks ~ treatment*days + treatment*gender,</pre>
      family = poisson,
       data = alc_dat_long,
       id = sid,
       corstr = "ar1")
summary(pois_gee_int) #ns
##
## Call:
  geeglm(formula = drinks ~ treatment * days + treatment * gender,
       family = poisson, data = alc_dat_long, id = sid, corstr = "ar1")
##
##
  Coefficients:
##
                           Estimate Std.err
                                                Wald Pr(>|W|)
## (Intercept)
                             5.2287 0.0178 85835.74
                                                       <2e-16 ***
                                                         0.38
## treatment2
                            -0.0197 0.0225
                                                0.76
                                                1.47
                                                         0.23
## treatment3
                            -0.0298 0.0247
## days30
                            -0.3939 0.0142
                                              764.23
                                                       <2e-16 ***
## days60
                            -0.4135 0.0125 1089.27
                                                       <2e-16 ***
## genderFemale
                            -1.0222 0.0235
                                             1884.16
                                                       <2e-16 ***
## treatment2:days30
                            -0.0129 0.0190
                                                0.46
                                                         0.50
## treatment3:days30
                            -0.3923 0.0209
                                              353.36
                                                       <2e-16 ***
                                              416.54
                                                       <2e-16 ***
## treatment2:days60
                            -0.3862 0.0189
                            -0.3875 0.0193
                                              403.80
                                                       <2e-16 ***
## treatment3:days60
## treatment2:genderFemale
                             0.0150 0.0351
                                                0.18
                                                         0.67
## treatment3:genderFemale
                             0.0159 0.0370
                                                0.18
                                                         0.67
```

```
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Estimated Scale Parameters:
              Estimate Std.err
                   1.9 0.112
## (Intercept)
## Correlation: Structure = ar1 Link = identity
##
## Estimated Correlation Parameters:
        Estimate Std.err
           0.549
                   0.033
## alpha
                             Maximum cluster size: 3
## Number of clusters:
                        314
pois_gee_intd <- geeglm(drinks ~ treatment*days + days*gender,</pre>
      family = poisson,
      data = alc_dat_long,
      id = sid,
      corstr = "ar1")
broom::tidy(pois_gee_intd, exponentiate = TRUE, conf.int = TRUE) #significant effect for day 60 but ver
## # A tibble: 12 x 7
##
     term
                      estimate std.error statistic p.value conf.low conf.high
##
      <chr>
                        <dbl>
                                 <dbl>
                                            <dbl> <dbl>
                                                             <dbl>
                                                                       <dbl>
## 1 (Intercept)
                                 0.0158
                                         1.09e+5 0
                                                           180.
                                                                    191.
                       185.
## 2 treatment2
                        0.985
                                 0.0191 6.17e-1 0.432
                                                             0.949
                                                                      1.02
## 3 treatment3
                        0.975
                                 0.0204
                                         1.51e+0 0.219
                                                                      1.02
                                                             0.937
## 4 days30
                        0.677
                                 0.0150
                                          6.72e+2 0
                                                             0.658
                                                                      0.698
## 5 days60
                        0.670
                                 0.0136
                                         8.62e+2 0
                                                             0.652
                                                                      0.688
## 6 genderFemale
                       0.368
                               0.0172
                                         3.38e+3 0
                                                             0.356
                                                                      0.381
                                         5.39e-1 0.463
## 7 treatment2:days~
                        0.986
                                 0.0190
                                                             0.950
                                                                      1.02
## 8 treatment3:days~
                        0.675
                                 0.0209
                                          3.53e+2 0
                                                             0.648
                                                                      0.703
## 9 treatment2:days~
                        0.677
                                 0.0189
                                          4.24e+2 0
                                                             0.653
                                                                      0.703
## 10 treatment3:days~
                        0.677
                                 0.0193
                                         4.08e+2 0
                                                             0.652
                                                                      0.703
                                          6.55e-1 0.418
                                                             0.947
                                                                      1.02
## 11 days30:genderFe~
                         0.984
                                 0.0199
                                                                       0.991
## 12 days60:genderFe~
                         0.953
                                 0.0197
                                          5.84e+0 0.0156
                                                             0.917
anova(pois_gee, pois_gee_int)
## Analysis of 'Wald statistic' Table
## Model 1 drinks ~ treatment * days + treatment * gender
## Model 2 drinks ~ treatment * days + gender
   Df
          X2 P(>|Chi|)
## 1 2 0.257
                  0.88
anova(pois_gee, pois_gee_intd)
## Analysis of 'Wald statistic' Table
##
## Model 1 drinks ~ treatment * days + days * gender
## Model 2 drinks ~ treatment * days + gender
##
   Df
       X2 P(>|Chi|)
## 1 2 5.86
                0.053 .
```

```
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##Trajectories##
#get contrast of treatment 3 days 60 to treatment 3 days 30
#get contrast of treatment 1 days 60 to treatment 1 days 30
#get contrast of treatment 3 days 60 to treatment 3 days 30
## Treatment effect ##
#get contrast of treatment 3 days 60 to treatment 2 days 60
#get contrast of treatment 3 days 60 to treatment 2 days 20
colnames(pois_gee$geese$vbeta) = names(coef(pois_gee))
rownames(pois_gee$geese$vbeta) = names(coef(pois_gee))
var_b0 = pois_gee$geese$vbeta["(Intercept)","(Intercept)"]
var_b1 = pois_gee$geese$vbeta["treatment2","treatment2"]
var_b2 = pois_gee$geese$vbeta["(Intercept)","(Intercept)"]
var_b3 = pois_gee$geese$vbeta["(Intercept)","(Intercept)"]
var_b4 = pois_gee$geese$vbeta["(Intercept)","(Intercept)"]
var_b6 = pois_gee$geese$vbeta["genderFemale","genderFemale"]
cov_b0b6 = pois_gee$geese$vbeta["genderFemale","(Intercept)"]
women_count_baseline = exp(coef(pois_gee)["(Intercept)"] + coef(pois_gee)["genderFemale"])
women_var_baseline = var_b0 + var_b6 + 2*cov_b0b6
women_ci_baseline = c(exp(coef(pois_gee)["(Intercept)"] + coef(pois_gee)["genderFemale"] - qchisq(0.975
var_b4 = pois_gee$geese$vbeta["days30","days30"]
var_b5 = pois_gee$geese$vbeta["days60","days60"]
cov_b4b5 = pois_gee$geese$vbeta["days30","days60"]
d30_to_d60_rr = exp(coef(pois_gee)["days60"] - coef(pois_gee)["days30"])
d30_d60_var = var_b4 + var_b5 - 2*cov_b4b5
d30_{to}d60_{ci} = c(exp(coef(pois_gee)["days60"] - coef(pois_gee)["days30"] - qchisq(0.975, 1)*sqrt(d30_d)
#treatment 2 to treatment 1
var_trt2 = pois_gee$geese$vbeta["treatment2","treatment2"]
var_trt2d30 = pois_gee$geese$vbeta["treatment2:days30","treatment2:days30"]
cov_day30trt2 = pois_gee$geese$vbeta["treatment2:days30","treatment2"]
trt2day30totrt1day30_rr = exp(coef(pois_gee)["treatment2"] + coef(pois_gee)["treatment2:days30"])
trt2day30totrt1day30_var = var_trt2 + var_trt2d30 - 2*cov_day30trt2
trt2day30totrt1day30_ci = c(exp(log(trt2day30totrt1day30_rr) - qchisq(0.975, 1)*sqrt(trt2day30totrt1day
var_trt2d60 = pois_gee$geese$vbeta["treatment2:days60","treatment2:days60"]
cov_day60trt2 = pois_gee$geese$vbeta["treatment2:days60","treatment2"]
trt2day60totrt1day60_rr = exp(coef(pois_gee)["treatment2"] + coef(pois_gee)["treatment2:days60"])
trt2day60totrt1day60_var = var_trt2 + var_trt2d60 + 2*cov_day60trt2
trt2day60totrt1day60_ci = c(exp(log(trt2day60totrt1day60_rr) - qchisq(0.975, 1)*sqrt(trt2day60totrt1day
cov_day60day30trt2 = pois_gee$geese$vbeta["treatment2:days60","treatment2:days30"]
trt2trt1_day60today30_r_rr = exp(coef(pois_gee)["treatment2:days60"] - coef(pois_gee)["treatment2:days3
trt2trt1_day60today30_r_var = var_trt2d30 + var_trt2d60 - 2*cov_day60day30trt2
trt2trt1_day60today30_r_ci = c(exp(log(trt2trt1_day60today30_r_rr) - qchisq(0.975, 1)*sqrt(trt2trt1_day
#treatment 3 to treatment 1
var_trt3 = pois_gee$geese$vbeta["treatment3","treatment3"]
```

```
var_trt3d30 = pois_gee$geese$vbeta["treatment3:days30","treatment3:days30"]
cov_day30trt3 = pois_gee$geese$vbeta["treatment3:days30","treatment3"]
trt3day30totrt1day30_rr = exp(coef(pois_gee)["treatment3"] + coef(pois_gee)["treatment3:days30"])
trt3day30totrt1day30_var = var_trt3 + var_trt3d30 - 2*cov_day30trt3
trt3day30totrt1day30_ci = c(exp(log(trt3day30totrt1day30_rr) - qchisq(0.975, 1)*sqrt(trt3day30totrt1day
var_trt3d60 = pois_gee$geese$vbeta["treatment3:days60","treatment3:days60"]
cov_day60trt3 = pois_gee$geese$vbeta["treatment3:days60","treatment3"]
trt3day60totrt1day60_rr = exp(coef(pois_gee)["treatment3"] + coef(pois_gee)["treatment3:days60"])
trt3day60totrt1day60_var = var_trt3 + var_trt3d60 + 2*cov_day60trt3
trt3day60totrt1day60_ci = c(exp(log(trt3day60totrt1day60_rr) - qchisq(0.975, 1)*sqrt(trt3day60totrt1day
cov_day60day30trt3 = pois_gee$geee$vbeta["treatment3:days60","treatment3:days30"]
trt3trt1_day60today30_r_rr = exp(coef(pois_gee)["treatment3:days60"] - coef(pois_gee)["treatment3:days3
trt3trt1_day60today30_r_var = var_trt3d30 + var_trt3d60 - 2*cov_day60day30trt3
trt3trt1_day60today30_r_ci = c(exp(log(trt3trt1_day60today30_r_rr) - qchisq(0.975, 1)*sqrt(trt3trt1_day
#treatment 3 to treatment 2
cov_day30trt3trt2 = pois_gee$geese$vbeta["treatment3:days30","treatment2"]
cov_day30trt3day30trt2 = pois_gee$geese$vbeta["treatment3:days30","treatment2:days30"]
cov_trt3day30trt2 = pois_gee$geese$vbeta["treatment3","treatment2:days30"]
cov_trt3trt2 = pois_gee$geese$vbeta["treatment3","treatment2"]
trt3day30totrt2day30_var = var_trt3 + var_trt3d30 + var_trt2 + var_trt2d30 + 2*cov_day30trt3 + 2*cov_day
trt3day30totrt2day30_ci = c(exp(log(trt3day30totrt2day30_rr) - qchisq(0.975, 1)*sqrt(trt3day30totrt2day
cov_day60trt3trt2 = pois_gee$geese$vbeta["treatment3:days60","treatment2"]
cov_day60trt3day60trt2 = pois_gee$geese$vbeta["treatment3:days60","treatment2:days60"]
cov_trt3day60trt2 = pois_gee$geese$vbeta["treatment3","treatment2:days60"]
cov_trt3trt2 = pois_gee$geese$vbeta["treatment3","treatment2"]
trt3day60totrt2day60_rr = exp(coef(pois_gee)["treatment3"] + coef(pois_gee)["treatment3:days60"] -
                               coef(pois_gee)["treatment2"] - coef(pois_gee)["treatment2:days60"])
trt3day60totrt2day60_var = var_trt3 + var_trt3d30 + var_trt2 + var_trt2d30 + 2*cov_day30trt3 + 2*cov_day
trt3day60totrt2day60_ci = c(exp(log(trt3day60totrt2day60_rr) - qchisq(0.975, 1)*sqrt(trt3day60totrt2day
broom::tidy(pois_gee, exponentiate = TRUE, conf.int = TRUE) %>%
  dplyr::select(term, estimate, std.error, conf.low, conf.high) %>%
  knitr::kable(caption = "Ratio of expected counts from Poisson GEE model (Exchangeable covariance)")
```

Table 3: Ratio of expected counts from Poisson GEE model (Exchangeable covariance)

term	estimate	std.error	conf.low	conf.high
(Intercept)	185.870	0.016	180.253	191.662
treatment2	0.984	0.019	0.948	1.022
treatment3	0.975	0.020	0.937	1.014
days30	0.674	0.014	0.656	0.694
days60	0.661	0.013	0.645	0.678
genderFemale	0.365	0.014	0.354	0.375

term	estimate	std.error	conf.low	conf.high
treatment2:days30	0.987	0.019	0.951	1.025
treatment3:days30	0.676	0.021	0.648	0.704
treatment2:days60	0.680	0.019	0.655	0.705
treatment3:days60	0.679	0.019	0.654	0.705