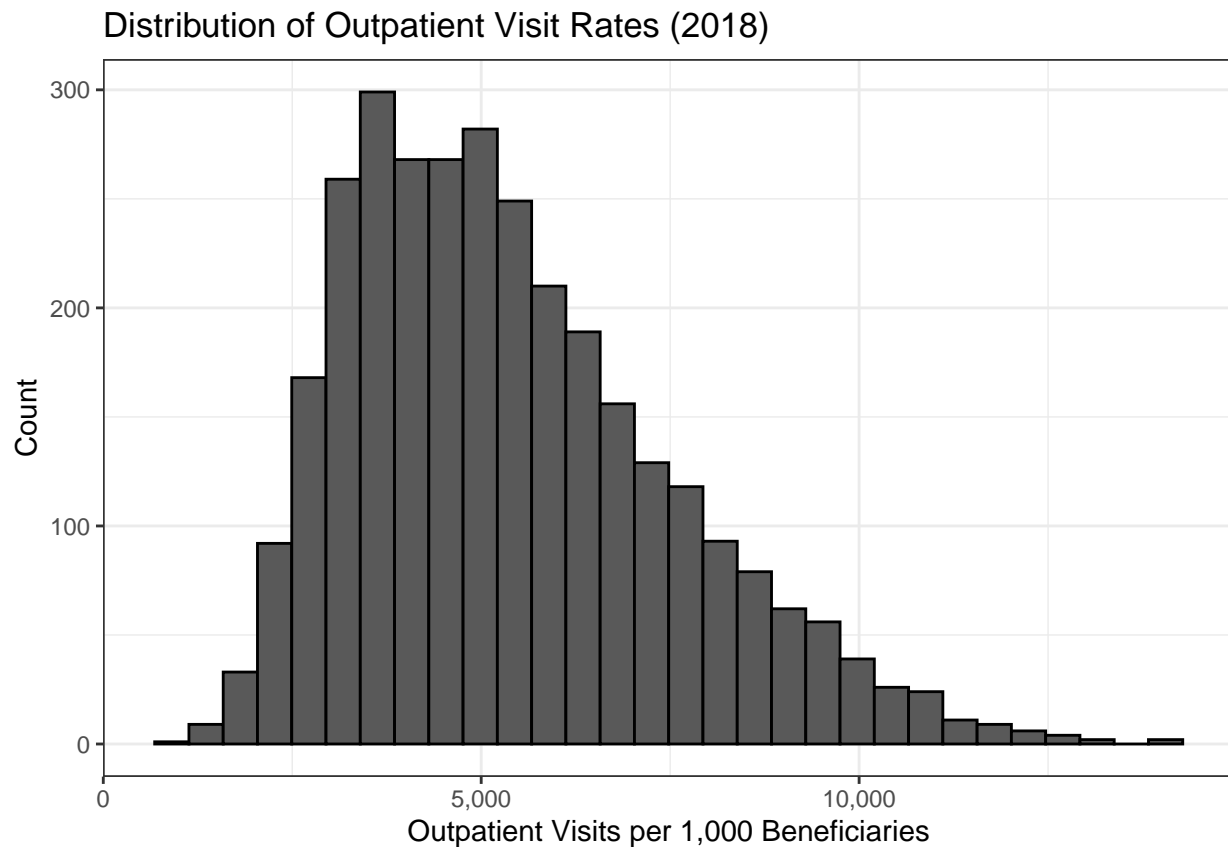


Section 5: Outpatient Visits Model

We built a similar model to the emergency department visits model to model the relationship between outpatient visits per 1,000 beneficiaries and Medicaid Expansion status. We added potential covariates one by one and evaluated the fit using likelihood ratio tests, with the intention of getting the best fit possible with this data. We did not identify any of these covariates as potentially confounding the relationship between the outpatient visit rate and Medicaid expansion status.

Check distribution of outcome

```
medicare %>%  
  ggplot(aes(x= op_visits_per_1000_beneficiaries)) +  
  geom_histogram(color = "black") +  
  scale_x_continuous(labels = scales::comma) +  
  scale_y_continuous(labels = scales::comma) +  
  xlab("Outpatient Visits per 1,000 Beneficiaries") +  
  ylab("Count") +  
  ggtitle("Distribution of Outpatient Visit Rates (2018)") +  
  theme_bw()  
  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

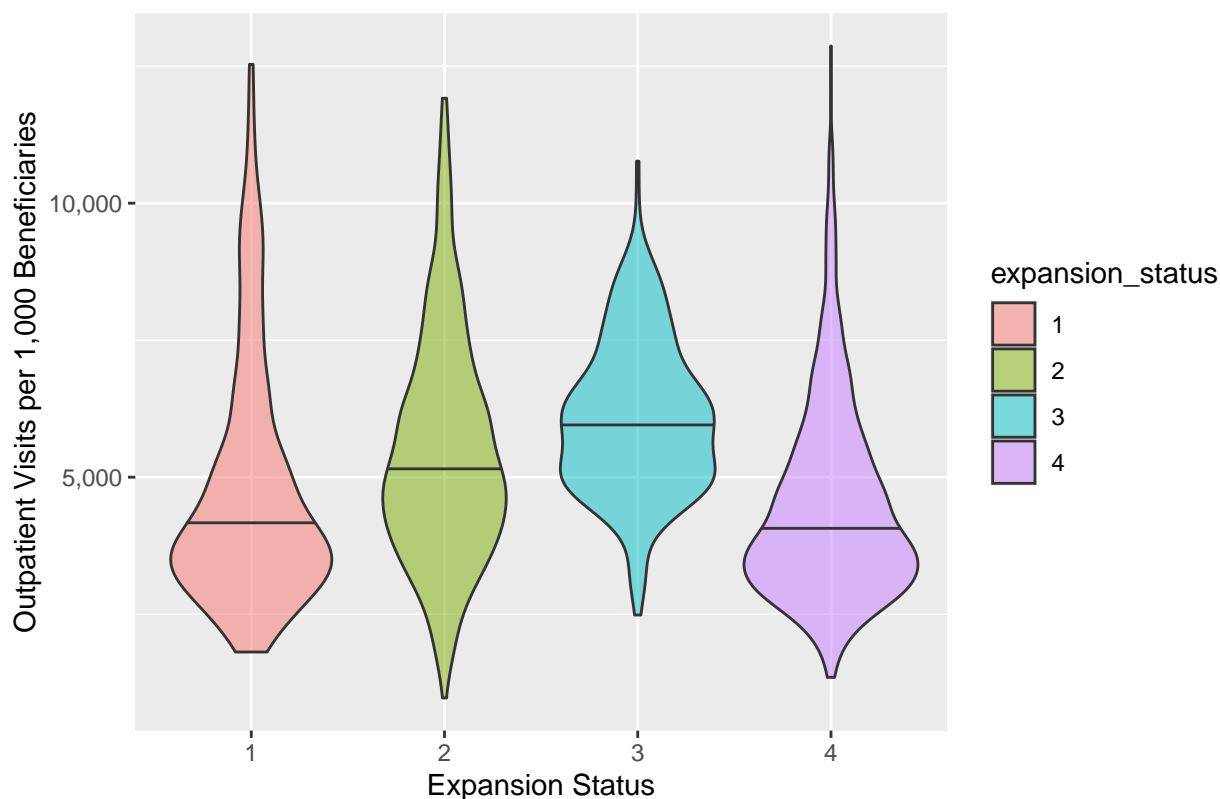


The outcome is pretty right skewed.

Create figure for paper:

```
medicare %>%
  na.omit() %>%
  ggplot(aes(x = expansion_status, y = op_visits_per_1000_beneficiaries, fill = expansion_status)) +
  geom_violin(alpha = 0.5, draw_quantiles = c(0.5)) +
  xlab("Expansion Status") +
  ylab("Outpatient Visits per 1,000 Beneficiaries") +
  scale_y_continuous(labels = scales::comma) +
  ggtitle("2018 Average Outpatient Visits per 1,000 Beneficiaries")
```

2018 Average Outpatient Visits per 1,000 Beneficiaries



We next check missingness in the outcome

```
medicare %>%
  select(op_visits_per_1000_beneficiaries) %>%
  summarize_all(.funs = function(x)
    return(paste0(round(sum(is.na(x))/length(x), 3)*100, "%")) %>%
  t() %>% data.frame()
```

```
##
## op_visits_per_1000_beneficiaries 0%
```

Model Fitting

First we fit the base model:

```
mod.nb_op<- MASS::glm.nb( op_visits_per_1000_beneficiaries ~ expansion_status + offset(log(year)), data=medicare)
summary(mod.nb_op)
```

```
##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
##   offset(log(year)), data = medicare, init.theta = 6.935188947,
##   link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6128  -0.8389  -0.2005   0.5423   3.3233
##
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.036689   0.026034  39.820 < 2e-16 ***
## expansion_status2 0.005551   0.029005   0.191   0.848
## expansion_status3 0.082962   0.032214   2.575   0.010 *
## expansion_status4 -0.135574   0.027672  -4.899 9.61e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(6.9352) family taken to be 1)
##
## Null deviance: 3368.9  on 3138  degrees of freedom
## Residual deviance: 3214.2  on 3135  degrees of freedom
## (5 observations deleted due to missingness)
## AIC: 56444
##
## Number of Fisher Scoring iterations: 1
##
##              Theta:  6.935
##             Std. Err.:  0.171
##
## 2 x log-likelihood:  -56434.396
# Model diagnostics table
model_diagnostics <- data.frame(
  Model = "OP Visits ~ Exp",
  AIC = AIC(mod.nb_op),
  BIC = BIC(mod.nb_op)
)
```

Improve Model Fit Using Other Covariates The covariates percent_male, percent_eligible_for_medicare, average_age, do not satisfy the criteria for confounding the relationship between Medicaid Expansion status and Outpatient Visit rates, so we will add each of these covariates one by one to determine if they improve the model fit using Likelihood Ratio tests.

Percent Male

```
mod.nb_op_c1 <- MASS::glm.nb(op_visits_per_1000_beneficiaries ~ expansion_status + percent_male + offset(log(year)), data = medicare, init.theta = 6.937525053,
summary(mod.nb_op_c1)

##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
##   percent_male + offset(log(year)), data = medicare, init.theta = 6.937525053,
##   link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5894  -0.8414  -0.1972   0.5482   3.2941
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.174738   0.147657   7.956 1.78e-15 ***
## expansion_status2 0.003099   0.029025   0.107  0.9150
```

```
## expansion_status3 0.080705 0.032238 2.503 0.0123 *
## expansion_status4 -0.138889 0.027800 -4.996 5.85e-07 ***
## percent_male -0.289999 0.307542 -0.943 0.3457
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(6.9375) family taken to be 1)
##
## Null deviance: 3368.3 on 3137 degrees of freedom
## Residual deviance: 3213.1 on 3133 degrees of freedom
## (6 observations deleted due to missingness)
## AIC: 56428
##
## Number of Fisher Scoring iterations: 1
##
##
## Theta: 6.938
## Std. Err.: 0.171
##
## 2 x log-likelihood: -56416.249
# Use Likelihood Ratio test to determine if confounder is necessary
anova(mod.nb_op, mod.nb_op_c1) # Yes

## Likelihood ratio tests of Negative Binomial Models
##
## Response: op_visits_per_1000_beneficiaries
##
## Model theta Resid. df
## 1 expansion_status + offset(log(year)) 6.935189 3135
## 2 expansion_status + percent_male + offset(log(year)) 6.937525 3133
## 2 x log-lik. Test df LR stat. Pr(Chi)
## 1 -56434.40
## 2 -56416.25 1 vs 2 2 18.14667 0.0001146832

# # Model diagnostics table
model_diagnostics <- model_diagnostics %>%
  bind_rows(
    data.frame(
      Model = "OP Visits ~ Exp + Male",
      AIC = AIC(mod.nb_op_c1),
      BIC = BIC(mod.nb_op_c1)
    )
  )
)
```

By the LRT, we don't add percent male into the model.

Percent Eligible for Medicaid

```
mod.nb_op_c2 <- MASS::glm.nb(op_visits_per_1000_beneficiaries ~ expansion_status + percent_male + percent_eligible_for_medicaid + offset(log(year)),
summary(mod.nb_op_c2)

##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
## percent_male + percent_eligible_for_medicaid + offset(log(year)),
## data = medicare, init.theta = 7.020015036, link = log)
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6895  -0.8519  -0.1910   0.5472   3.2012
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.166474   0.147815   7.891 2.99e-15 ***
## expansion_status2    0.003537   0.028863   0.123  0.90246
## expansion_status3    0.082627   0.032065   2.577  0.00997 **
## expansion_status4   -0.147864   0.027697  -5.339 9.36e-08 ***
## percent_male      -0.035246   0.309107  -0.114  0.90922
## percent_eligible_for_medicare -0.521759   0.078639  -6.635 3.25e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(7.02) family taken to be 1)
##
##      Null deviance: 3402.6  on 3128  degrees of freedom
## Residual deviance: 3203.0  on 3123  degrees of freedom
##      (15 observations deleted due to missingness)
## AIC: 56232
##
## Number of Fisher Scoring iterations: 1
##
##              Theta:  7.020
##              Std. Err.:  0.174
##
## 2 x log-likelihood:  -56218.163
##
# Use Anova to determine if confounder is necessary
anova(mod.nb_op_c1, mod.nb_op_c2) # Yes

## Likelihood ratio tests of Negative Binomial Models
##
## Response: op_visits_per_1000_beneficiaries
##
## 1 expansion_status + percent_male + offset(log(year))
## 2 expansion_status + percent_male + percent_eligible_for_medicare + offset(log(year))
##      theta Resid. df    2 x log-lik.    Test    df LR stat. Pr(Chi)
## 1 6.937525    3133    -56416.25
## 2 7.020015    3123    -56218.16 1 vs 2    10 198.086      0

## Model diagnostics table
model_diagnostics <- model_diagnostics %>%
  bind_rows(
    data.frame(
      Model = "OP Visits ~ Exp + Male + % Elig",
      AIC = AIC(mod.nb_op_c2),
      BIC = BIC(mod.nb_op_c2)
    )
  )
```

By the LRT, we add percent eligible for medicare to the model.

Average Age

```
mod.nb_op_c3 <- MASS::glm.nb(op_visits_per_1000_beneficiaries ~ expansion_status + percent_male + percent_eligible_for_medicaid + average_age + offset(log(year)), data = medicare, init.theta = 7.454848469, link = log)
summary(mod.nb_op_c3)
```

```
##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
##   percent_male + percent_eligible_for_medicaid + average_age +
##   offset(log(year)), data = medicare, init.theta = 7.454848469,
##   link = log)
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.8435  -0.7984  -0.1663   0.5399   3.5085
##
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -5.0419510   0.4539884 -11.106 < 2e-16 ***
## expansion_status2    0.0003124   0.0280161   0.011  0.99110
## expansion_status3    0.0930664   0.0311390   2.989  0.00280 **
## expansion_status4   -0.1525237   0.0268853  -5.673 1.40e-08 ***
## percent_male        0.9245799   0.3062435   3.019  0.00254 **
## percent_eligible_for_medicaid 0.5704108   0.1034070   5.516 3.46e-08 ***
## average_age         0.0775238   0.0054210  14.301 < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## (Dispersion parameter for Negative Binomial(7.4548) family taken to be 1)
```

```
##      Null deviance: 3613.0  on 3128  degrees of freedom
## Residual deviance: 3198.7  on 3122  degrees of freedom
## (15 observations deleted due to missingness)
## AIC: 56038
```

```
## Number of Fisher Scoring iterations: 1
```

```
##              Theta:  7.455
##            Std. Err.: 0.185
##
```

```
## 2 x log-likelihood:  -56021.788
```

```
# Use Likelihood ratio test to determine if covariate is necessary
```

```
anova(mod.nb_op_c2, mod.nb_op_c3) # Yes
```

```
## Likelihood ratio tests of Negative Binomial Models
```

```
##
```

```
## Response: op_visits_per_1000_beneficiaries
```

```
##
```

```
## 1 expansion_status + percent_male + percent_eligible_for_medicaid + offset(log(year))
```

```
## 2 expansion_status + percent_male + percent_eligible_for_medicaid + average_age + offset(log(year))
```

```
##      theta Resid. df    2 x log-lik.  Test      df LR stat. Pr(Chi)
```

```
## 1 7.020015    3123    -56218.16
```

```
## 2 7.454848    3122    -56021.79 1 vs 2      1 196.3753      0
```

```
## Model diagnostics table
model_diagnostics <- model_diagnostics %>%
  bind_rows(
    data.frame(
      Model = "OP Visits ~ Exp + Male + % Elig + Age",
      AIC = AIC(mod.nb_op_c3),
      BIC = BIC(mod.nb_op_c3)
    )
  )
)
```

By the LRT, we add average age to the model.

```
mod.nb_op_1 <- MASS::glm.nb(op_visits_per_1000_beneficiaries ~ expansion_status + percent_male + percent_eligible_for_medicaid + average_age + percent_eligible_for_medicaid_2 + offset(log(year)), data = medicare %>%
  mutate(percent_eligible_for_medicaid_2 = percent_eligible_for_medicaid^2),
  init.theta = 7.472198409, link = log)

summary(mod.nb_op_1)
```

Check for quadratic effects of percent_eligible_for_medicaid

```
##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
##   percent_male + percent_eligible_for_medicaid + average_age +
##   percent_eligible_for_medicaid_2 + offset(log(year)), data = medicare %>%
##   mutate(percent_eligible_for_medicaid_2 = percent_eligible_for_medicaid^2),
##   init.theta = 7.472198409, link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.7429  -0.7895  -0.1682   0.5449   3.4648
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -5.478734   0.482297  -11.360 < 2e-16 ***
## expansion_status2  0.004834   0.028016   0.173 0.863012
## expansion_status3  0.096449   0.031115   3.100 0.001937 **
## expansion_status4 -0.145427   0.026940  -5.398 6.73e-08 ***
## percent_male     1.047697   0.308984   3.391 0.000697 ***
## percent_eligible_for_medicaid 1.412240   0.311833   4.529 5.93e-06 ***
## average_age      0.081432   0.005626  14.475 < 2e-16 ***
## percent_eligible_for_medicaid_2 -1.568580   0.543490  -2.886 0.003900 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(7.4722) family taken to be 1)
##
##      Null deviance: 3621.4  on 3128  degrees of freedom
## Residual deviance: 3198.6  on 3121  degrees of freedom
## (15 observations deleted due to missingness)
## AIC: 56032
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta: 7.472
```



```
##          Std. Err.: 0.185
##
## 2 x log-likelihood: -56014.206
anova(mod.nb_op_c3, mod.nb_op_1) # Yes

## Likelihood ratio tests of Negative Binomial Models
##
## Response: op_visits_per_1000_beneficiaries
##
## 1 expansion_status + percent_male + percent_eligible_for_medicare +
## 2 expansion_status + percent_male + percent_eligible_for_medicare + average_age + percent_eligible_for_medicare^2
##      theta Resid. df    2 x log-lik. Test    df LR stat.    Pr(Chi)
## 1 7.454848    3122    -56021.79
## 2 7.472198    3121    -56014.21 1 vs 2      1 7.581868 0.005895833

model_diagnostics <- model_diagnostics %>%
  bind_rows(
    data.frame(
      Model = "OP Visits ~ Exp + Male + % Elig + Age + % Elig^2",
      AIC = AIC(mod.nb_op_1),
      BIC = BIC(mod.nb_op_1)
    )
  )
)
```

Effect Modification We think there may be effect modification between % eligible for medicaid and expansion status, since in general Medicaid expansion means a higher percentage of the population qualifies for Medicaid.

```
mod.nb_op_2 <- MASS::glm.nb(op_visits_per_1000_beneficiaries ~ expansion_status + percent_eligible_for_medicare + percent_eligible_for_medicare^2, data = medicare %>%
  mutate(percent_eligible_for_medicare_2 = percent_eligible_for_medicare^2),
  init.theta = 7.513451751, link = log)

summary(mod.nb_op_2)

##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
##   percent_eligible_for_medicare + percent_male + average_age +
##   percent_eligible_for_medicare_2 + percent_eligible_for_medicare *
##   expansion_status + offset(log(year)), data = medicare %>%
##   mutate(percent_eligible_for_medicare_2 = percent_eligible_for_medicare^2),
##   init.theta = 7.513451751, link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6585  -0.7972  -0.1636   0.5411   3.5233
##
## Coefficients:
##              Estimate Std. Error z value
## (Intercept)    -5.474332   0.484969 -11.288
## expansion_status2 -0.175608   0.079954  -2.196
## expansion_status3 -0.024012   0.089130  -0.269
## expansion_status4 -0.180317   0.076553  -2.355
## percent_eligible_for_medicare  1.107559   0.435865   2.541
## percent_male      0.913908   0.309579   2.952
## average_age      0.083272   0.005652  14.732
## percent_eligible_for_medicare_2 -1.652752   0.545039  -3.032
```

```
## expansion_status2:percent_eligible_for_medicaid 0.827468 0.342228 2.418
## expansion_status3:percent_eligible_for_medicaid 0.546809 0.382664 1.429
## expansion_status4:percent_eligible_for_medicaid 0.122890 0.329460 0.373
##                                     Pr(>|z|)
## (Intercept) < 2e-16 ***
## expansion_status2 0.02807 *
## expansion_status3 0.78762
## expansion_status4 0.01850 *
## percent_eligible_for_medicaid 0.01105 *
## percent_male 0.00316 **
## average_age < 2e-16 ***
## percent_eligible_for_medicaid_2 0.00243 **
## expansion_status2:percent_eligible_for_medicaid 0.01561 *
## expansion_status3:percent_eligible_for_medicaid 0.15302
## expansion_status4:percent_eligible_for_medicaid 0.70915
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(7.5135) family taken to be 1)
##
## Null deviance: 3641.4 on 3128 degrees of freedom
## Residual deviance: 3198.2 on 3118 degrees of freedom
## (15 observations deleted due to missingness)
## AIC: 56020
##
## Number of Fisher Scoring iterations: 1
##
##
## Theta: 7.513
## Std. Err.: 0.186
##
## 2 x log-likelihood: -55996.250
```

```
anova(mod.nb_op_1, mod.nb_op_2)
```

```
## Likelihood ratio tests of Negative Binomial Models
```

```
##
```

```
## Response: op_visits_per_1000_beneficiaries
```

```
##
```

```
## 1 expansion_status + percent_male + percent_eligible_for_medicaid
```

```
## 2 expansion_status + percent_eligible_for_medicaid + percent_male + average_age + percent_eligible_for_medicaid^2
```

```
##      theta Resid. df    2 x log-lik. Test      df LR stat.      Pr(Chi)
```

```
## 1 7.472198    3121    -56014.21
```

```
## 2 7.513452    3118    -55996.25 1 vs 2      3 17.9561 0.0004491143
```

```
model_diagnostics <- model_diagnostics %>%
```

```
  bind_rows(
```

```
    data.frame(
```

```
      Model = "OP Visits ~ Exp + Male + % Elig + Age + % Elig^2 + % Elig*Exp",
```

```
      AIC = AIC(mod.nb_op_2),
```

```
      BIC = BIC(mod.nb_op_2)
```

```
    )
```

```
)
```

By the LRT, we do add the interaction between Medicaid expansion status and percent_eligible_for_medicaid.

```
model_diagnostics
```

Model Diagnostics

```
##                                     Model      AIC
## 1                                     OP Visits ~ Exp 56444.40
## 2                                     OP Visits ~ Exp + Male 56428.25
## 3                                     OP Visits ~ Exp + Male + % Elig 56232.16
## 4                                     OP Visits ~ Exp + Male + % Elig + Age 56037.79
## 5                                     OP Visits ~ Exp + Male + % Elig + Age + % Elig^2 56032.21
## 6 OP Visits ~ Exp + Male + % Elig + Age + % Elig^2 + % Elig*Exp 56020.25
##      BIC
## 1 56474.65
## 2 56464.56
## 3 56274.50
## 4 56086.18
## 5 56086.64
## 6 56092.83
```

We have the smallest AIC and BIC with the full model.

```
mod.pois_ed <- glm(op_visits_per_1000_beneficiaries ~ expansion_status + percent_eligible_for_medicaid
# Evaluate overdispersion
deviance(mod.pois_ed)/mod.pois_ed$df.residual # 710.5393
```

Check Poisson model

```
## [1] 710.5393
pearson.stat_ed <- sum((na.omit(medicare %>% select(percent_male, percent_eligible_for_medicaid, average_
pearson.stat_ed/mod.pois_ed$df.residual # 730.8859

## [1] 730.8859
# Evidence of overdispersion
mean(medicare$op_visits_per_1000_beneficiaries, na.rm = T)

## [1] 5380.558
var(medicare$op_visits_per_1000_beneficiaries, na.rm = T)

## [1] 4521870
```

Due to the evidence of dispersion, the negative binomial model seems to be the appropriate appropriate model.

Interpretation Get estimated IRR and 95% confidence interval for expansion status with the effects of percent eligible for Medicaid (2 v 1)

```
exp(coef(mod.nb_op_2)[2] + coef(mod.nb_op_2)[9])

## expansion_status2
##      1.919107
exp((coef(mod.nb_op_2)[2] + coef(mod.nb_op_2)[9]) + c(-1,1)*1.96*(vcov(mod.nb_op_2)[2,2] + vcov(mod.nb_op_2)[9,9]))

## [1] 1.362446 2.703206
```

Get estimated IRR and 95% confidence interval for expansion status with the effects of percent eligible for Medicaid (3 v 1)

```
exp(coef(mod.nb_op_2)[3] + coef(mod.nb_op_2)[10])
```

```
## expansion_status3  
## 1.68674
```

```
exp((coef(mod.nb_op_2)[3] + coef(mod.nb_op_2)[10]) + c(-1,1)*1.96*(vcov(mod.nb_op_2)[3,3] + vcov(mod.nb_op_2)[10,10]))
```

```
## [1] 1.099540 2.587528
```

Get estimated IRR and 95% confidence interval for expansion status with the effects of percent eligible for Medicaid (4 v 1)

```
exp(coef(mod.nb_op_2)[4] + coef(mod.nb_op_2)[11])
```

```
## expansion_status4  
## 0.9441903
```

```
exp((coef(mod.nb_op_2)[4] + coef(mod.nb_op_2)[11]) + c(-1,1)*1.96*(vcov(mod.nb_op_2)[4,4] + vcov(mod.nb_op_2)[11,11]))
```

```
## [1] 0.6878298 1.2960987
```