# Section 2: Fitting PQI Models

**PQIs of Interest**

For these models, we will be looking at modeling the relationship between various PQIs at the HRR (hospital-referral region) level and Medicaid expansion status. For model selection, we started with the base model and added in potential covariates one by one from our list of potential covariates that did not have significant missingness in order to improve the model fit. We using likelihood ratio tests at each step to determine if adding the covariate was an appropriate addition to the model (checking the reduced model against the full model).

**Exploratory analysis on covariates**

First we quantify the missingness in the outcomes:

```
tbl <- hrr %>%
  select( `pqi05_copd_or_asthma_in_older_adults_admission_rate_(age_40-64)`, `pqi05_copd_or_asthma_in_o
  summarize_all(.funs = function(x) return(paste0(round(sum(is.na(x))/length(x), 3)*100, "%"))) %>% t()

names(tbl) <- c("% of Missing values")
kableExtra::kable(tbl)
```

|                                                                    | % of Missing values |
|--------------------------------------------------------------------|---------------------|
| pqi05_copd_or_asthma_in_older_adults_admission_rate_(age_40-64)     | 1.3%                |
| pqi05_copd_or_asthma_in_older_adults_admission_rate_(age_65-74)     | 0.6%                |
| pqi05_copd_or_asthma_in_older_adults_admission_rate_(age_75+)       | 0.6%                |
| pqi07_hypertension_admission_rate_(age_<_65)                       | 22.1%               |
| pqi07_hypertension_admission_rate_(age_65-74)                      | 17.9%               |
| pqi07_hypertension_admission_rate_(age_75+)                       | 6.8%                |
| pqi08_chf_admission_rate_(age_65-74)                              | 0.6%                |
| pqi08_chf_admission_rate_(age_75+)                               | 0.6%                |
| pqi08_chf_admission_rate_(age_<_65)                              | 1%                  |

We have a very small percentage of missing data in our target covariates and 2 of the 3 PQIs. We believe PQI05 and PQI08 will be reasonable metrics to evaluate outpatient quality and it's relation to Medicaid expansion.

**Check distribution of outcomes**

```
hrr %>%
  filter(!is.na(expansion_status)) %>%
  select(year, `pqi05_copd_or_asthma_in_older_adults_admission_rate_(age_40-64)`, `pqi05_copd_or_asthma
  gather(age, pqi05, -year) %>%
  mutate(
    age = str_sub(age, -10)
  ) %>% ggplot(aes( x = pqi05)) + geom_histogram(color = "black") +
  scale_x_continuous(labels = scales::comma) +
  scale_y_continuous(labels = scales::comma) +
  xlab("PQI 05") +
  ylab("Count") +
```
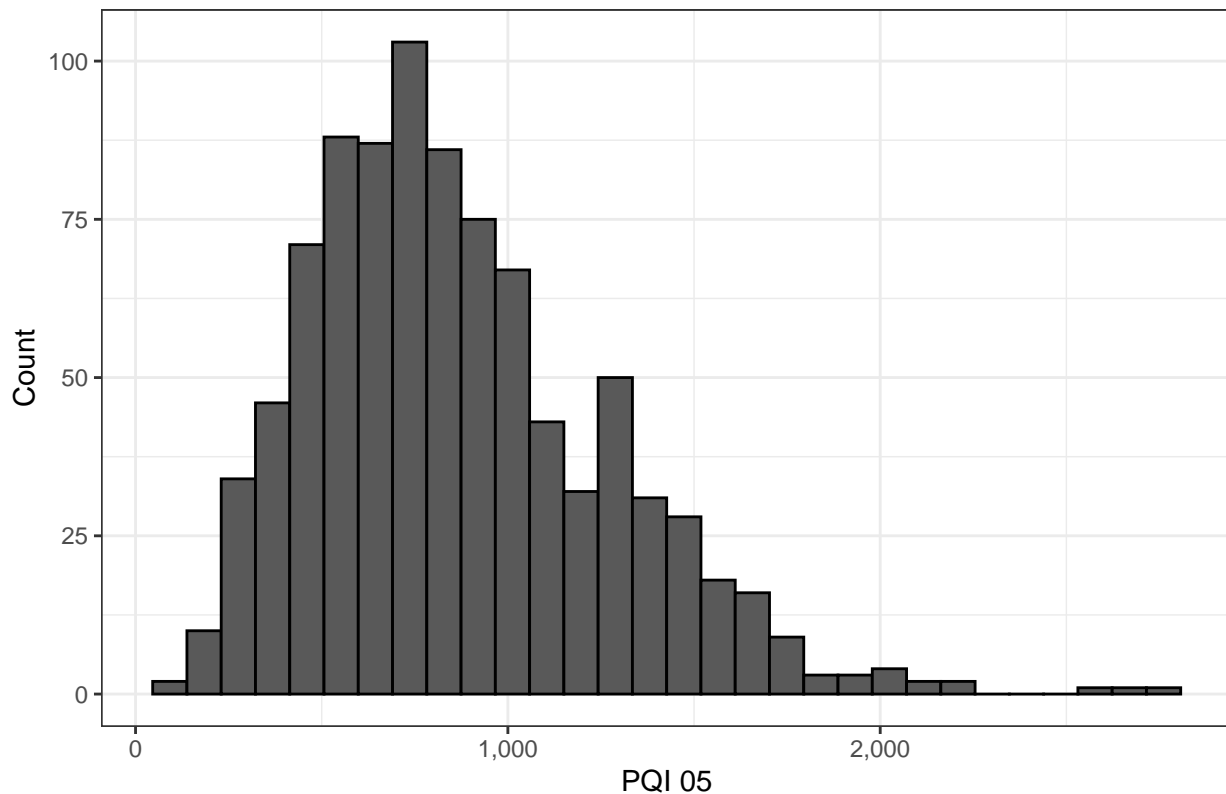
```
  ggtitle("Distribution of PQI 05: COPD/Asthma in Older Adults Admissions Rate (2007-2018)") +
  theme_bw()
```

**PQI05**

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 2 rows containing non-finite values (stat_bin).
```

Distribution of PQI 05: COPD/Asthma in Older Adults Admissions Rate (20



The distribution has a slight right skewe.

```
hrr %>%
  filter(!is.na(expansion_status)) %>%
  select(year, `pqi08_chf_admission_rate_(age_65-74)`, `pqi08_chf_admission_rate_(age_75+)`, `pqi08_chf
  gather(age, pqi08, -year) %>%
  mutate(
    age = str_sub(age, -10)
  ) %>% ggplot(aes( x = pqi08)) + geom_histogram(color = "black") +
  scale_x_continuous(labels = scales::comma) +
  scale_y_continuous(labels = scales::comma) +
  xlab("PQI 08") +
  ylab("Count") +
  ggtitle("Distribution of PQI 05: Congestive Heart Failure Admissions Rate (2007-2018)") +
  theme_bw()
```

**PQI08**

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

### Distribution of PQI 05: Congestive Heart Failure Admissions Rate (2007–20
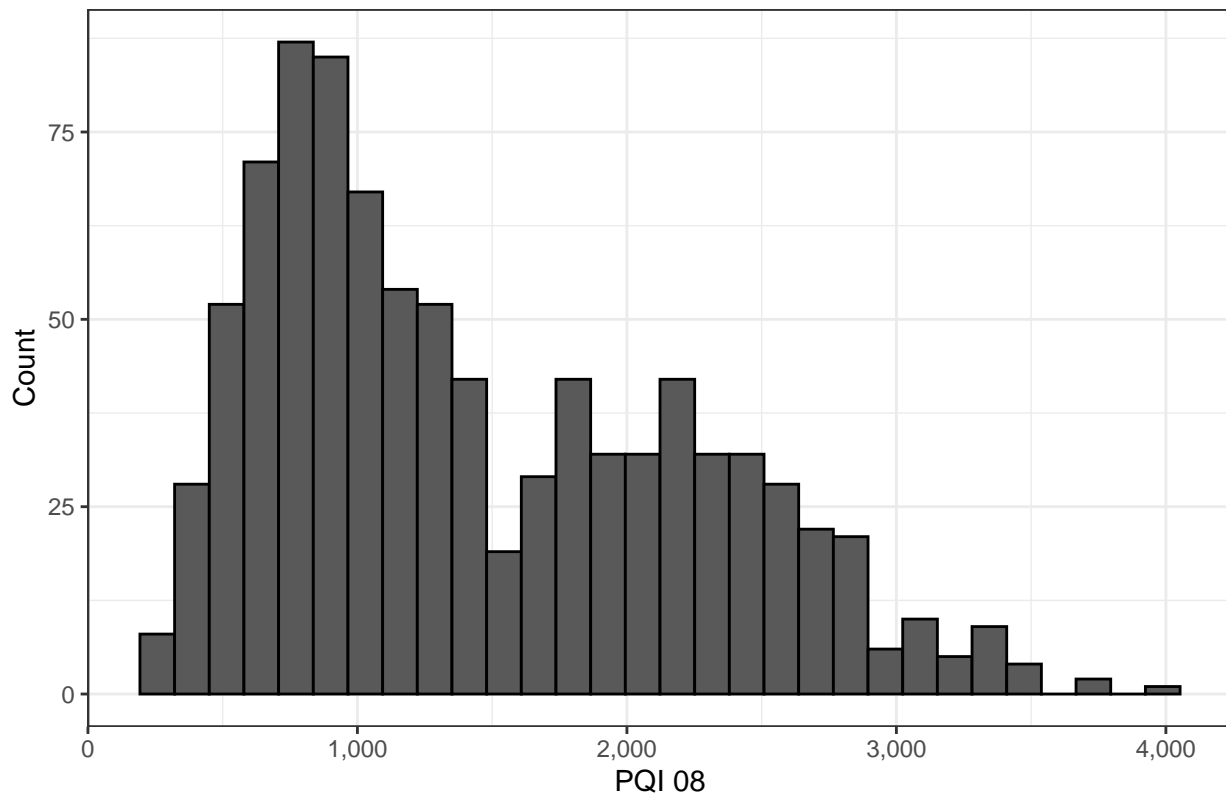


This outcome is almost bi-modally normally distributed (or just has a large right skew).

### Model Fitting

**PQI05**   Since these variables are rates, we will want to use Poisson models. There may be overdispersion, so I'm going to start fitting a negative binomial and then check if the poisson version has overdispersion.

```
pqi05_data <- hrr %>%
  filter(!is.na(expansion_status)) %>%
  select(year, hrr, state, expansion_status,  percent_male, percent_african_american, `percent_hispanic
  gather(age, pqi05, c("pqi05_copd_or_asthma_in_older_adults_admission_rate_(age_40-64)", "pqi05_copd_o
  mutate(age = str_extract(age, "(?<=age_).+(?=\\))"))
```

```
mod.nb_pqi05 <- MASS::glm.nb(pqi05 ~ `expansion_status` + offset(log(year)), data = pqi05_data)
summary(mod.nb_pqi05)
```

#### Start with the base model

```
##
## Call:
## MASS::glm.nb(formula = pqi05 ~ expansion_status + offset(log(year)),
##     data = pqi05_data, init.theta = 5.197820128, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.5765  -0.7998  -0.1908   0.5561   3.0121
```

```
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -1.19097    0.03805 -31.299  < 2e-16 ***
## expansion_status2  0.37661    0.04682   8.044 8.72e-16 ***
## expansion_status3  0.44497    0.05140   8.657  < 2e-16 ***
## expansion_status4  0.38405    0.04461   8.609  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(5.1978) family taken to be 1)
##
##     Null deviance: 1028.72  on 912  degrees of freedom
## Residual deviance:  942.22  on 909  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 13314
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  5.198
##           Std. Err.:  0.238
##
##  2 x log-likelihood:  -13304.003
```

We'll start adding in potential covariates and use likelihood ratio tests to determine if they should be added to the model.

```
mod.nb_pqi05_c1 <- MASS::glm.nb(pqi05 ~ `expansion_status` +  percent_male + offset(log(year)), data = 
summary(mod.nb_pqi05_c1)
```

**Percent Male**

```
##
## Call:
## MASS::glm.nb(formula = pqi05 ~ expansion_status + percent_male +
##     offset(log(year)), data = pqi05_data, init.theta = 5.485463972,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.6564  -0.7822  -0.1444   0.5367   3.0758
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        2.28027    0.46638   4.889 1.01e-06 ***
## expansion_status2  0.36410    0.04562   7.982 1.44e-15 ***
## expansion_status3  0.39757    0.05029   7.906 2.67e-15 ***
## expansion_status4  0.36599    0.04348   8.417  < 2e-16 ***
## percent_male      -7.57507    1.01389  -7.471 7.94e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(5.4855) family taken to be 1)
```

```
##
##     Null deviance: 1085.21  on 912  degrees of freedom
## Residual deviance:  940.77  on 908  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 13264
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  5.485
##          Std. Err.:  0.251
##
##  2 x log-likelihood:  -13252.224
```

```
anova(mod.nb_pqi05, mod.nb_pqi05_c1)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi05
##                                            Model    theta Resid. df
## 1            expansion_status + offset(log(year)) 5.197820       909
## 2 expansion_status + percent_male + offset(log(year)) 5.485464       908
##     2 x log-lik.   Test    df LR stat.     Pr(Chi)
## 1        -13304.00
## 2        -13252.22 1 vs 2     1 51.77882 6.211698e-13
```

By the LRT, we add percent_male into the model.

```
mod.nb_pqi05_c2 <- MASS::glm.nb(pqi05 ~ `expansion_status` +  percent_male + percent_african_american +
summary(mod.nb_pqi05_c2)
```

**Percent African American**

```
##
## Call:
## MASS::glm.nb(formula = pqi05 ~ expansion_status + percent_male +
##     percent_african_american + offset(log(year)), data = pqi05_data,
##     init.theta = 5.567089558, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.5948  -0.8405  -0.1495   0.5058   3.1249
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)               1.44437    0.51041   2.830 0.004658 **
## expansion_status2         0.36225    0.04529   7.998 1.27e-15 ***
## expansion_status3         0.38344    0.05019   7.641 2.16e-14 ***
## expansion_status4         0.33001    0.04418   7.469 8.08e-14 ***
## percent_male             -5.83219    1.10243  -5.290 1.22e-07 ***
## percent_african_american  0.71312    0.18361   3.884 0.000103 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(5.5671) family taken to be 1)
```

```
## 
##     Null deviance: 1101.23  on 912  degrees of freedom
## Residual deviance:  940.37  on 907  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 13252
## 
## Number of Fisher Scoring iterations: 1
## 
## 
##              Theta:  5.567
##          Std. Err.:  0.255
## 
##  2 x log-likelihood:  -13238.041
```

```r
anova(mod.nb_pqi05_c1, mod.nb_pqi05_c2)
```

```
## Likelihood ratio tests of Negative Binomial Models
## 
## Response: pqi05
##                                                                       Model
## 1                              expansion_status + percent_male + offset(log(year))
## 2 expansion_status + percent_male + percent_african_american + offset(log(year))
##      theta Resid. df   2 x log-lik.   Test   df LR stat.     Pr(Chi)
## 1 5.485464       908      -13252.22
## 2 5.567090       907      -13238.04 1 vs 2    1 14.18301 0.000165861
```

By the LRT, we add in percent African American into the model.

```r
mod.nb_pqi05_c3 <- MASS::glm.nb(pqi05 ~ `expansion_status` +  percent_male + percent_african_american +
summary(mod.nb_pqi05_c3)
```

**Percent Hispanic**

```
## 
## Call:
## MASS::glm.nb(formula = pqi05 ~ expansion_status + percent_male +
##     percent_african_american + percent_hispanic + offset(log(year)),
##     data = pqi05_data, init.theta = 5.591227791, link = log)
## 
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.6061  -0.8156  -0.1640   0.5104   3.1390
## 
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)                1.43579    0.50938   2.819 0.004822 **
## expansion_status2          0.33745    0.04647   7.262 3.82e-13 ***
## expansion_status3          0.35339    0.05204   6.791 1.11e-11 ***
## expansion_status4          0.31299    0.04475   6.994 2.67e-12 ***
## percent_male              -5.72874    1.10156  -5.201 1.99e-07 ***
## percent_african_american   0.69132    0.18350   3.767 0.000165 ***
## percent_hispanic          -0.32974    0.16179  -2.038 0.041540 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
```

```
## (Dispersion parameter for Negative Binomial(5.5912) family taken to be 1)
##
##     Null deviance: 1105.97  on 912  degrees of freedom
## Residual deviance:  940.26  on 906  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 13250
##
## Number of Fisher Scoring iterations: 1
##
##
##             Theta:  5.591
##         Std. Err.:  0.256
##
##  2 x log-likelihood:  -13233.900
```

```r
anova(mod.nb_pqi05_c2, mod.nb_pqi05_c3)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi05
##                                                                                         Model
## 1                   expansion_status + percent_male + percent_african_american + offset(log(year))
## 2 expansion_status + percent_male + percent_african_american + percent_hispanic + offset(log(year))
##      theta Resid. df   2 x log-lik.   Test   df LR stat.    Pr(Chi)
## 1 5.567090       907      -13238.04
## 2 5.591228       906      -13233.90 1 vs 2    1 4.141419 0.04184607
```

By the LRT, we add in percent hispanic into the model.

**Age category**   We add in the age category by which the data was originally split up.

```r
mod.nb_pqi05_c4 <- MASS::glm.nb(pqi05 ~ `expansion_status` +  percent_male + percent_african_american +
summary(mod.nb_pqi05_c4)
```

```
##
## Call:
## MASS::glm.nb(formula = pqi05 ~ expansion_status + percent_male +
##     percent_african_american + percent_hispanic + age + offset(log(year)),
##     data = pqi05_data, init.theta = 9.312393762, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8996  -0.7410  -0.0816   0.5543   3.3060
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 1.35219    0.39607   3.414  0.00064 ***
## expansion_status2           0.33926    0.03612   9.392  < 2e-16 ***
## expansion_status3           0.35952    0.04044   8.889  < 2e-16 ***
## expansion_status4           0.31244    0.03479   8.981  < 2e-16 ***
## percent_male               -4.95142    0.85590  -5.785 7.25e-09 ***
## percent_african_american    0.79997    0.14250   5.614 1.98e-08 ***
## percent_hispanic           -0.30207    0.12575  -2.402  0.01630 *
## age65-74                   -0.67622    0.02674 -25.284  < 2e-16 ***
## age75+                     -0.28164    0.02671 -10.545  < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(9.3124) family taken to be 1)
##
##      Null deviance: 1832.48  on 912  degrees of freedom
## Residual deviance:  930.39  on 904  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 12772
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  9.312
##           Std. Err.:  0.434
##
##  2 x log-likelihood:  -12752.317
```

```r
anova(mod.nb_pqi05_c3, mod.nb_pqi05_c4)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi05
##
## 1        expansion_status + percent_male + percent_african_american + percent_hispanic + offset(log(y
## 2 expansion_status + percent_male + percent_african_american + percent_hispanic + age + offset(log(y
##      theta Resid. df    2 x log-lik.    Test    df LR stat. Pr(Chi)
## 1 5.591228        906        -13233.90
## 2 9.312394        904        -12752.32 1 vs 2      2  481.583        0
```

By the LRT, we add age category into the model.

```r
mod.nb_pqi05_c5 <- MASS::glm.nb(pqi05 ~ `expansion_status` +  percent_male + percent_african_american +
summary(mod.nb_pqi05_c5)
```

**Percent Eligible for Medicaid**

```
##
## Call:
## MASS::glm.nb(formula = pqi05 ~ expansion_status + percent_male +
##     percent_african_american + percent_hispanic + age + percent_eligible_for_medicaid +
##     offset(log(year)), data = pqi05_data, init.theta = 10.56586819,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8638  -0.7618  -0.0617   0.5606   3.7466
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)                2.12508    0.37606   5.651 1.60e-08 ***
## expansion_status2          0.38198    0.03429  11.141  < 2e-16 ***
## expansion_status3          0.35968    0.03801   9.463  < 2e-16 ***
## expansion_status4          0.41442    0.03390  12.223  < 2e-16 ***
## percent_male              -7.48105    0.82676  -9.049  < 2e-16 ***
## percent_african_american   0.13527    0.14682   0.921    0.357
```

```
## percent_hispanic              -0.92578    0.13065  -7.086 1.38e-12 ***
## age65-74                       -0.68677    0.02513 -27.327  < 2e-16 ***
## age75+                         -0.28341    0.02509 -11.294  < 2e-16 ***
## percent_eligible_for_medicaid  2.06973    0.18317  11.299  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(10.5659) family taken to be 1)
##
##      Null deviance: 2075.52  on 912  degrees of freedom
## Residual deviance:  928.39  on 903  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 12657
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  10.566
##           Std. Err.:  0.495
##
##  2 x log-likelihood:  -12634.718
```

```r
anova(mod.nb_pqi05_c4, mod.nb_pqi05_c5)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi05
##
## 1                                 expansion_status + percent_male + percent_african_american + percen
## 2 expansion_status + percent_male + percent_african_american + percent_hispanic + age + percent_elig
##      theta Resid. df   2 x log-lik.    Test    df LR stat. Pr(Chi)
## 1  9.312394       904      -12752.32
## 2 10.565868       903      -12634.72 1 vs 2     1 117.5989       0
```

By the LRT, we add in the percent eligible for Medicaid.

```r
mod.nb_pqi05_1 <- MASS::glm.nb(pqi05 ~ `expansion_status` + percent_male + percent_african_american + p
summary(mod.nb_pqi05_1)
```

**Check for quadratic effects of `percent_eligible_for_medicaid`**

```
##
## Call:
## MASS::glm.nb(formula = pqi05 ~ expansion_status + percent_male +
##     percent_african_american + percent_hispanic + age + percent_eligible_for_medicaid +
##     percent_eligible_for_medicaid_2 + offset(log(year)), data = pqi05_data %>%
##     mutate(percent_eligible_for_medicaid_2 = percent_eligible_for_medicaid^2),
##     init.theta = 10.73197305, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.7449  -0.7610  -0.0724   0.5409   3.9483
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
```

9

```
## (Intercept)                         1.97470    0.37547   5.259 1.45e-07 ***
## expansion_status2                    0.39179    0.03413  11.479  < 2e-16 ***
## expansion_status3                    0.35426    0.03773   9.390  < 2e-16 ***
## expansion_status4                    0.41901    0.03368  12.439  < 2e-16 ***
## percent_male                        -7.79114    0.82379  -9.458  < 2e-16 ***
## percent_african_american            0.13263    0.14571   0.910    0.363
## percent_hispanic                    -0.68909    0.14343  -4.804 1.55e-06 ***
## age65-74                            -0.68874    0.02494 -27.617  < 2e-16 ***
## age75+                              -0.28437    0.02490 -11.421  < 2e-16 ***
## percent_eligible_for_medicaid        4.79309    0.71814   6.674 2.48e-11 ***
## percent_eligible_for_medicaid_2     -6.01438    1.53259  -3.924 8.70e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(10.732) family taken to be 1)
##
##     Null deviance: 2107.66  on 912  degrees of freedom
## Residual deviance:  928.16  on 902  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 12644
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  10.732
##           Std. Err.:  0.503
##
##  2 x log-likelihood:  -12620.248
```

```
anova(mod.nb_pqi05_c4, mod.nb_pqi05_1)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi05
##
## 1                                                                 expansion_status + percent_male +
## 2 expansion_status + percent_male + percent_african_american + percent_hispanic + age + percent_elig
##       theta Resid. df   2 x log-lik.    Test    df LR stat. Pr(Chi)
## 1  9.312394       904       -12752.32
## 2 10.731973       902       -12620.25 1 vs 2     2 132.0694       0
```

By the LRT, we add the quadratic effects of the percent eligible for medicaid.

**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_pqi05_2 <-MASS::glm.nb(pqi05 ~ `expansion_status` +  percent_male + percent_african_american + p
summary(mod.nb_pqi05_2 )
```

```
##
## Call:
## MASS::glm.nb(formula = pqi05 ~ expansion_status + percent_male +
##     percent_african_american + percent_hispanic + age + percent_eligible_for_medicaid +
##     percent_eligible_for_medicaid_2 + expansion_status * percent_eligible_for_medicaid +
##     offset(log(year)), data = pqi05_data %>% mutate(percent_eligible_for_medicaid_2 = percent_eligibl
```

```
##     init.theta = 10.81176474, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8050  -0.7623  -0.0844   0.5308   4.1183
##
## Coefficients:
##                                                  Estimate Std. Error z value
## (Intercept)                                       1.97190    0.37866    5.208
## expansion_status2                                 0.32494    0.10302    3.154
## expansion_status3                                 0.56754    0.13497    4.205
## expansion_status4                                 0.48624    0.09793    4.965
## percent_male                                     -7.80840    0.83033   -9.404
## percent_african_american                          0.20430    0.15024    1.360
## percent_hispanic                                 -0.65931    0.14620   -4.510
## age65-74                                         -0.68823    0.02485  -27.697
## age75+                                           -0.28327    0.02481  -11.418
## percent_eligible_for_medicaid                     4.72213    0.85160    5.545
## percent_eligible_for_medicaid_2                  -5.69466    1.57719   -3.611
## expansion_status2:percent_eligible_for_medicaid  0.37762    0.44829    0.842
## expansion_status3:percent_eligible_for_medicaid -0.98168    0.59522   -1.649
## expansion_status4:percent_eligible_for_medicaid -0.35142    0.41896   -0.839
##                                                  Pr(>|z|)
## (Intercept)                                      1.91e-07 ***
## expansion_status2                                0.001610 **
## expansion_status3                                2.61e-05 ***
## expansion_status4                                6.85e-07 ***
## percent_male                                      < 2e-16 ***
## percent_african_american                         0.173880
## percent_hispanic                                 6.50e-06 ***
## age65-74                                          < 2e-16 ***
## age75+                                            < 2e-16 ***
## percent_eligible_for_medicaid                    2.94e-08 ***
## percent_eligible_for_medicaid_2                  0.000305 ***
## expansion_status2:percent_eligible_for_medicaid 0.399586
## expansion_status3:percent_eligible_for_medicaid 0.099092 .
## expansion_status4:percent_eligible_for_medicaid 0.401593
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(10.8118) family taken to be 1)
##
##     Null deviance: 2123.10  on 912  degrees of freedom
## Residual deviance:  928.07  on 899  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 12643
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  10.812
##           Std. Err.:  0.507
##
##  2 x log-likelihood:  -12613.388
```

```r
anova(mod.nb_pqi05_1, mod.nb_pqi05_2)
```

```
## Likelihood ratio tests of Negative Binomial Models
## 
## Response: pqi05
## 
## 1                                                 expansion_status + percent_male + percent_africa
## 2 expansion_status + percent_male + percent_african_american + percent_hispanic + age + percent_elig
##       theta Resid. df   2 x log-lik.   Test   df LR stat.   Pr(Chi)
## 1 10.73197       902        -12620.25
## 2 10.81176       899        -12613.39 1 vs 2    3 6.859423 0.07651599
```

By the LRT, we do not add the interaction term to the model.

```r
AIC(mod.nb_pqi05)
```

**Model Diagnostics**

```
## [1] 13314
```

```r
AIC(mod.nb_pqi05_c1)
```

```
## [1] 13264.22
```

```r
AIC(mod.nb_pqi05_c2)
```

```
## [1] 13252.04
```

```r
AIC(mod.nb_pqi05_c3)
```

```
## [1] 13249.9
```

```r
AIC(mod.nb_pqi05_c4)
```

```
## [1] 12772.32
```

```r
AIC(mod.nb_pqi05_c5)
```

```
## [1] 12656.72
```

```r
AIC(mod.nb_pqi05_1)
```

```
## [1] 12644.25
```

The full model has the lowest AIC.

```r
BIC(mod.nb_pqi05)
```

```
## [1] 13338.09
```

```r
BIC(mod.nb_pqi05_c1)
```

```
## [1] 13293.12
```

```r
BIC(mod.nb_pqi05_c2)
```

```
## [1] 13285.76
```

```r
BIC(mod.nb_pqi05_c3)
```

```
## [1] 13288.43
```

```r
BIC(mod.nb_pqi05_c4)
```

```
## [1] 12820.48
```

```r
BIC(mod.nb_pqi05_c5)
```

```
## [1] 12709.7
```

```r
BIC(mod.nb_pqi05_1)
```

```
## [1] 12702.05
```

The full model has the lowest AIC and BIC of all the models.

**Check Poisson model**   As a last step, we'll check the model as a Poisson model, and check if there is truly overdispersion.

```r
mod.pois_pqi05 <- glm(pqi05 ~ expansion_status + percent_male +
    percent_african_american + percent_hispanic + age + percent_eligible_for_medicaid +
    percent_eligible_for_medicaid_2 +
    offset(log(year)), data = pqi05_data %>% mutate(percent_eligible_for_medicaid_2 = percent_eligible_

# Evaluate overdispersion
deviance(mod.pois_pqi05)/mod.pois_pqi05$df.residual #  78.17053
```

```
## [1] 78.17053
```

```r
pearson.stat_pqi05 <- sum((na.omit(pqi05_data)$pqi05 - fitted(mod.pois_pqi05))^2/fitted(mod.pois_pqi05))
pearson.stat_pqi05/mod.pois_pqi05$df.residual # 79.35596
```

```
## [1] 79.35596
```

```r
# Evidence of overdispersion
mean(pqi05_data$pqi05, na.rm = T) # 866.4775
```

```
## [1] 866.4775
```

```r
var(pqi05_data$pqi05, na.rm = T) # 157324.5
```

```
## [1] 157324.5
```

By these, we can say the negative binomial is the right choice.

```r
exp(coef(mod.nb_pqi05_1))
```

**Interpretation**

```
##                     (Intercept)                expansion_status2
##                    7.204430e+00                     1.479632e+00
##               expansion_status3                expansion_status4
##                    1.425125e+00                     1.520460e+00
##                    percent_male        percent_african_american
##                    4.133815e-04                     1.141825e+00
##                percent_hispanic                          age65-74
##                    5.020330e-01                     5.022104e-01
##                           age75+    percent_eligible_for_medicaid
##                    7.524900e-01                     1.206742e+02
## percent_eligible_for_medicaid_2
##                    2.443356e-03
```

13

```
exp(confint(mod.nb_pqi05_1))
```

```
## Waiting for profiling to be done...

##                                      2.5 %        97.5 %
## (Intercept)                   3.396255e+00 1.527924e+01
## expansion_status2             1.383489e+00 1.581847e+00
## expansion_status3             1.323337e+00 1.534566e+00
## expansion_status4             1.422925e+00 1.623995e+00
## percent_male                  7.882399e-05 2.173371e-03
## percent_african_american      8.501881e-01 1.536220e+00
## percent_hispanic              3.782871e-01 6.684109e-01
## age65-74                      4.782099e-01 5.274140e-01
## age75+                        7.165962e-01 7.901798e-01
## percent_eligible_for_medicaid   2.862236e+01 5.036960e+02
## percent_eligible_for_medicaid_2 1.155054e-04 5.291054e-02
```

**PQI08**   Since these variables are rates, we will want to use Poisson models. There may be overdispersion, so I'm going to start fitting a negative binomial and then check if the poisson version has overdispersion.

```
pqi08_data <- hrr %>%
  filter(!is.na(expansion_status)) %>%
  select(year, hrr, state, expansion_status,  percent_male, percent_african_american, `percent_hispanic
  gather(age, pqi08, c("pqi08_chf_admission_rate_(age_<_65)", "pqi08_chf_admission_rate_(age_65-74)", "
  mutate(age = str_extract(age, "(?<=age_).+(?=\\))"))
```

```
mod.nb_pqi08 <- MASS::glm.nb(pqi08 ~ `expansion_status` + offset(log(year)), data = pqi08_data)
summary(mod.nb_pqi08)
```

**Start with the base model**

```
##
## Call:
## MASS::glm.nb(formula = pqi08 ~ expansion_status + offset(log(year)),
##     data = pqi08_data, init.theta = 3.471413904, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6165  -0.9591  -0.3042   0.7107   2.0079
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.46835    0.04626 -10.125  < 2e-16 ***
## expansion_status2  0.10140    0.05701   1.779 0.075314 .
## expansion_status3  0.23732    0.06262   3.790 0.000151 ***
## expansion_status4  0.16100    0.05430   2.965 0.003026 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.4714) family taken to be 1)
##
##     Null deviance: 973.41  on 913  degrees of freedom
## Residual deviance: 957.34  on 910  degrees of freedom
##   (1 observation deleted due to missingness)
```

```
## AIC: 14579
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  3.471
##           Std. Err.:  0.156
##
##  2 x log-likelihood:  -14569.068
```

We'll start adding in potential covariates and use likelihood ratio tests to determine if they should be added
to the model.

```
mod.nb_pqi08_c1 <- MASS::glm.nb(pqi08 ~ `expansion_status` +  percent_male + offset(log(year)), data = p
summary(mod.nb_pqi08_c1)
```

**Percent Male**

```
##
## Call:
## MASS::glm.nb(formula = pqi08 ~ expansion_status + percent_male +
##     offset(log(year)), data = pqi08_data, init.theta = 3.600767565,
##     link = log)
##
## Deviance Residuals:
##     Min        1Q    Median        3Q       Max
## -2.4705   -0.9875   -0.3146    0.7153    2.0902
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       3.08815    0.57443   5.376 7.61e-08 ***
## expansion_status2 0.08506    0.05602   1.518  0.12889
## expansion_status3 0.18719    0.06179   3.029  0.00245 **
## expansion_status4 0.14328    0.05338   2.684  0.00727 **
## percent_male     -7.75935    1.24876  -6.214 5.18e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.6008) family taken to be 1)
##
##     Null deviance: 1009.57  on 913  degrees of freedom
## Residual deviance:  955.81  on 909  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 14545
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  3.601
##           Std. Err.:  0.162
##
##  2 x log-likelihood:  -14532.644
```

```
anova(mod.nb_pqi08, mod.nb_pqi08_c1)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi08
##                                               Model    theta Resid. df
## 1                expansion_status + offset(log(year)) 3.471414       910
## 2 expansion_status + percent_male + offset(log(year)) 3.600768       909
##     2 x log-lik.   Test   df LR stat.      Pr(Chi)
## 1       -14569.07
## 2       -14532.64 1 vs 2    1 36.42389 1.587457e-09
```

By the LRT, we add percent_male into the model.

```
mod.nb_pqi08_c2 <- MASS::glm.nb(pqi08 ~ `expansion_status` +  percent_male + percent_african_american +
summary(mod.nb_pqi08_c2)
```

**Percent African American**

```
##
## Call:
## MASS::glm.nb(formula = pqi08 ~ expansion_status + percent_male +
##     percent_african_american + offset(log(year)), data = pqi08_data,
##     init.theta = 3.769564704, link = log)
##
## Deviance Residuals:
##    Min       1Q   Median       3Q      Max
## -2.4891  -0.9775  -0.3492   0.6786   2.2747
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)                1.27611    0.61895   2.062   0.0392 *
## expansion_status2          0.07935    0.05477   1.449   0.1474
## expansion_status3          0.14624    0.06073   2.408   0.0160 *
## expansion_status4          0.06457    0.05343   1.209   0.2268
## percent_male              -3.98088    1.33688  -2.978   0.0029 **
## percent_african_american  1.53074    0.22272   6.873 6.28e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.7696) family taken to be 1)
##
##     Null deviance: 1056.75  on 913  degrees of freedom
## Residual deviance:  953.96  on 908  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 14501
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  3.770
##           Std. Err.:  0.170
##
##  2 x log-likelihood:  -14487.193
```

```
anova(mod.nb_pqi08_c1, mod.nb_pqi08_c2)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi08
##                                                               Model
## 1                        expansion_status + percent_male + offset(log(year))
## 2 expansion_status + percent_male + percent_african_american + offset(log(year))
##      theta Resid. df   2 x log-lik.   Test   df LR stat.     Pr(Chi)
## 1 3.600768       909      -14532.64
## 2 3.769565       908      -14487.19 1 vs 2    1 45.45124 1.564859e-11
```

By the LRT, we add in percent African American into the model.

```
mod.nb_pqi08_c3 <- MASS::glm.nb(pqi08 ~ `expansion_status` +  percent_male + percent_african_american +
summary(mod.nb_pqi08_c3)
```

**Percent Hispanic**

```
##
## Call:
## MASS::glm.nb(formula = pqi08 ~ expansion_status + percent_male +
##     percent_african_american + percent_hispanic + offset(log(year)),
##     data = pqi08_data, init.theta = 3.787438553, link = log)
##
## Deviance Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4963  -0.9818  -0.3646  0.6812  2.3433
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)                1.30097    0.61756   2.107  0.03515 *
## expansion_status2          0.10917    0.05621   1.942  0.05210 .
## expansion_status3          0.18193    0.06298   2.889  0.00387 **
## expansion_status4          0.08155    0.05412   1.507  0.13185
## percent_male              -4.13968    1.33555  -3.100  0.00194 **
## percent_african_american   1.56848    0.22254   7.048 1.81e-12 ***
## percent_hispanic           0.40554    0.19598   2.069  0.03852 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(3.7874) family taken to be 1)
##
##     Null deviance: 1061.74  on 913  degrees of freedom
## Residual deviance:  953.77  on 907  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 14499
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  3.787
##          Std. Err.:  0.170
##
```

```
##  2 x log-likelihood:  -14482.508
anova(mod.nb_pqi08_c2, mod.nb_pqi08_c3)

## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi08
##                                                                                      Model
## 1                      expansion_status + percent_male + percent_african_american + offset(log(year))
## 2 expansion_status + percent_male + percent_african_american + percent_hispanic + offset(log(year))
##      theta Resid. df    2 x log-lik.   Test    df LR stat.    Pr(Chi)
## 1 3.769565       908       -14487.19
## 2 3.787439       907       -14482.51 1 vs 2     1 4.684588 0.03043436
```

By the LRT, we add percent hispanic to the model.

```
mod.nb_pqi08_c4 <- MASS::glm.nb(pqi08 ~ `expansion_status` +  percent_male + percent_african_american +
summary(mod.nb_pqi08_c4)
```

**Age category**

```
##
## Call:
## MASS::glm.nb(formula = pqi08 ~ expansion_status + percent_male +
##      percent_african_american + percent_hispanic + age + offset(log(year)),
##      data = pqi08_data, init.theta = 13.74646014, link = log)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -3.8050   -0.6991   -0.0622    0.5877    3.4104
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 0.99357    0.32590   3.049 0.002298 **
## expansion_status2           0.11559    0.02965   3.898 9.70e-05 ***
## expansion_status3           0.19052    0.03322   5.735 9.73e-09 ***
## expansion_status4           0.10545    0.02855   3.694 0.000221 ***
## percent_male               -3.92248    0.70431  -5.569 2.56e-08 ***
## percent_african_american    1.87339    0.11722  15.981  < 2e-16 ***
## percent_hispanic            0.49103    0.10330   4.753 2.00e-06 ***
## age65-74                   -0.44704    0.02202 -20.301  < 2e-16 ***
## age75+                      0.64509    0.02196  29.382  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(13.7465) family taken to be 1)
##
##      Null deviance: 3821.51  on 913  degrees of freedom
## Residual deviance:  926.26  on 905  degrees of freedom
##    (1 observation deleted due to missingness)
## AIC: 13276
##
## Number of Fisher Scoring iterations: 1
##
##
```

```
##              Theta:   13.746
##          Std. Err.:   0.644
##
##  2 x log-likelihood:  -13256.279
```

```
anova(mod.nb_pqi08_c3, mod.nb_pqi08_c4)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi08
##                                                                                              I
## 1       expansion_status + percent_male + percent_african_american + percent_hispanic + offset(log(yo
## 2 expansion_status + percent_male + percent_african_american + percent_hispanic + age + offset(log(yo
##       theta Resid. df   2 x log-lik.   Test    df LR stat. Pr(Chi)
## 1  3.787439      907      -14482.51
## 2 13.746460      905      -13256.28 1 vs 2     2 1226.229        0
```

By the LRT, we add in age category to the model.

```
mod.nb_pqi08_c5 <- MASS::glm.nb(pqi08 ~ `expansion_status` +  percent_male + percent_african_american +
summary(mod.nb_pqi08_c5)
```

**Percent Eligible for Medicaid**

```
##
## Call:
## MASS::glm.nb(formula = pqi08 ~ expansion_status + percent_male +
##     percent_african_american + percent_hispanic + age + percent_eligible_for_medicaid +
##     offset(log(year)), data = pqi08_data, init.theta = 14.35255937,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.9043  -0.7056  -0.0705   0.5978   3.5575
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    1.32394    0.32235   4.107 4.01e-05 ***
## expansion_status2              0.14053    0.02931   4.794 1.63e-06 ***
## expansion_status3              0.19079    0.03252   5.867 4.44e-09 ***
## expansion_status4              0.15683    0.02897   5.413 6.20e-08 ***
## percent_male                  -5.05646    0.70870  -7.135 9.69e-13 ***
## percent_african_american       1.55864    0.12588  12.382  < 2e-16 ***
## percent_hispanic               0.18220    0.11186   1.629    0.103
## age65-74                      -0.45301    0.02156 -21.013  < 2e-16 ***
## age75+                         0.64174    0.02149  29.860  < 2e-16 ***
## percent_eligible_for_medicaid  1.01699    0.15702   6.477 9.36e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(14.3526) family taken to be 1)
##
##     Null deviance: 3987.99  on 913  degrees of freedom
## Residual deviance:  925.61  on 904  degrees of freedom
##   (1 observation deleted due to missingness)
```

```
## AIC: 13238
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  14.353
##          Std. Err.:  0.673
##
##  2 x log-likelihood:  -13216.242
```

```r
anova(mod.nb_pqi08_c4, mod.nb_pqi08_c5)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi08
##
## 1                               expansion_status + percent_male + percent_african_american + percer
## 2 expansion_status + percent_male + percent_african_american + percent_hispanic + age + percent_elig
##     theta Resid. df   2 x log-lik.   Test   df LR stat.      Pr(Chi)
## 1 13.74646      905       -13256.28
## 2 14.35256      904       -13216.24 1 vs 2    1 40.03666 2.492413e-10
```

By the LRT we add in percent eligible for medicaid into the model.

```r
mod.nb_pqi08_1 <- MASS::glm.nb(pqi08 ~ `expansion_status` +  percent_male + percent_african_american + p
summary(mod.nb_pqi08_1)
```

**Check for quadratic effects of `percent_eligible_for_medicaid`**

```
##
## Call:
## MASS::glm.nb(formula = pqi08 ~ expansion_status + percent_male +
##     percent_african_american + percent_hispanic + age + percent_eligible_for_medicaid +
##     percent_eligible_for_medicaid_2 + offset(log(year)), data = pqi08_data %>%
##     mutate(percent_eligible_for_medicaid_2 = percent_eligible_for_medicaid^2),
##     init.theta = 14.56522529, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.9464  -0.7028  -0.0576   0.6006   3.5345
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      1.20537    0.32198   3.744 0.000181 ***
## expansion_status2                0.14902    0.02920   5.104 3.33e-07 ***
## expansion_status3                0.18700    0.03229   5.791 7.00e-09 ***
## expansion_status4                0.16041    0.02880   5.569 2.56e-08 ***
## percent_male                    -5.32774    0.70645  -7.542 4.64e-14 ***
## percent_african_american         1.55893    0.12497  12.474  < 2e-16 ***
## percent_hispanic                 0.38696    0.12278   3.152 0.001623 **
## age65-74                        -0.45338    0.02140 -21.183  < 2e-16 ***
## age75+                           0.64182    0.02134  30.082  < 2e-16 ***
## percent_eligible_for_medicaid    3.26043    0.61531   5.299 1.17e-07 ***
## percent_eligible_for_medicaid_2 -4.94633    1.31252  -3.769 0.000164 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(14.5652) family taken to be 1)
##
##      Null deviance: 4046.37  on 913  degrees of freedom
## Residual deviance:  925.46  on 903  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 13227
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  14.565
##           Std. Err.:  0.683
##
##  2 x log-likelihood:  -13202.676
```

```r
anova(mod.nb_pqi08_c5, mod.nb_pqi08_1)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi08
##
## 1                                    expansion_status + percent_male + percent_african_american + per
## 2 expansion_status + percent_male + percent_african_american + percent_hispanic + age + percent_elig
##       theta Resid. df   2 x log-lik.   Test    df LR stat.      Pr(Chi)
## 1 14.35256       904       -13216.24
## 2 14.56523       903       -13202.68 1 vs 2     1  13.5666 0.000230246
```

By the LRT we add the quadratic effects of percent eligible for medicaid to the model.

**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.

```r
mod.nb_pqi08_2 <- MASS::glm.nb(pqi08 ~ `expansion_status` +  percent_male + percent_african_american + 
summary(mod.nb_pqi08_2)
```

```
##
## Call:
## MASS::glm.nb(formula = pqi08 ~ expansion_status + percent_male +
##     percent_african_american + percent_hispanic + age + percent_eligible_for_medicaid +
##     percent_eligible_for_medicaid_2 + expansion_status * percent_eligible_for_medicaid +
##     offset(log(year)), data = pqi08_data %>% mutate(percent_eligible_for_medicaid_2 = percent_eligibl
##     init.theta = 14.62256022, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.9506  -0.7084  -0.0503   0.6013   3.5055
##
## Coefficients:
##                                           Estimate Std. Error z value
## (Intercept)                                1.15321    0.32526   3.546
## expansion_status2                          0.25612    0.08836   2.899
## expansion_status3                          0.31206    0.11582   2.694
## expansion_status4                          0.30691    0.08396   3.655
```

```
## percent_male                                      -5.48210    0.71333  -7.685
## percent_african_american                           1.56698    0.12908  12.139
## percent_hispanic                                   0.42635    0.12539   3.400
## age65-74                                           -0.45247    0.02136 -21.181
## age75+                                              0.64285    0.02129  30.189
## percent_eligible_for_medicaid                       3.85819    0.73098   5.278
## percent_eligible_for_medicaid_2                    -5.32649    1.35333  -3.936
## expansion_status2:percent_eligible_for_medicaid -0.44823    0.38483  -1.165
## expansion_status3:percent_eligible_for_medicaid -0.53568    0.51099  -1.048
## expansion_status4:percent_eligible_for_medicaid -0.66679    0.35949  -1.855
##                                                  Pr(>|z|)
## (Intercept)                                      0.000392 ***
## expansion_status2                                0.003748 **
## expansion_status3                                0.007052 **
## expansion_status4                                0.000257 ***
## percent_male                                     1.53e-14 ***
## percent_african_american                          < 2e-16 ***
## percent_hispanic                                 0.000673 ***
## age65-74                                          < 2e-16 ***
## age75+                                            < 2e-16 ***
## percent_eligible_for_medicaid                    1.31e-07 ***
## percent_eligible_for_medicaid_2                  8.29e-05 ***
## expansion_status2:percent_eligible_for_medicaid 0.244125
## expansion_status3:percent_eligible_for_medicaid 0.294489
## expansion_status4:percent_eligible_for_medicaid 0.063620 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(14.6226) family taken to be 1)
##
##     Null deviance: 4062.10  on 913  degrees of freedom
## Residual deviance:  925.44  on 900  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 13229
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  14.623
##           Std. Err.:  0.686
##
##  2 x log-likelihood:  -13199.070
```

```
anova(mod.nb_pqi08_1, mod.nb_pqi08_2)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: pqi08
##
## 1                                              expansion_status + percent_male + percent_africa
## 2 expansion_status + percent_male + percent_african_american + percent_hispanic + age + percent_elig
##      theta Resid. df   2 x log-lik.   Test   df LR stat. Pr(Chi)
## 1 14.56523      903        -13202.68
## 2 14.62256      900        -13199.07 1 vs 2    3 3.605778  0.3073
```

By the LRT, we don't add the interaction term to the model.

```
AIC(mod.nb_pqi08)
```

**Model Diagnostics**

```
## [1] 14579.07
```

```
AIC(mod.nb_pqi08_c1)
```

```
## [1] 14544.64
```

```
AIC(mod.nb_pqi08_c2)
```

```
## [1] 14501.19
```

```
AIC(mod.nb_pqi08_c3)
```

```
## [1] 14498.51
```

```
AIC(mod.nb_pqi08_c4)
```

```
## [1] 13276.28
```

```
AIC(mod.nb_pqi08_c5)
```

```
## [1] 13238.24
```

```
AIC(mod.nb_pqi08_1)
```

```
## [1] 13226.68
```

```
BIC(mod.nb_pqi08)
```

```
## [1] 14603.16
```

```
BIC(mod.nb_pqi08_c1)
```

```
## [1] 14573.55
```

```
BIC(mod.nb_pqi08_c2)
```

```
## [1] 14534.92
```

```
BIC(mod.nb_pqi08_c3)
```

```
## [1] 14537.05
```

```
BIC(mod.nb_pqi08_c4)
```

```
## [1] 13324.46
```

```
BIC(mod.nb_pqi08_c5)
```

```
## [1] 13291.24
```

```
BIC(mod.nb_pqi08_1)
```

```
## [1] 13284.49
```

The full model has the lowest AIC and BIC of all the models.

```r
mod.pois_pqi08 <- glm(pqi08 ~ `expansion_status` + `percent_male` + `percent_african_american` + percen
```

```r
# Evaluate overdispersion
deviance(mod.pois_pqi08)/mod.pois_pqi08$df.residual #  90.85094
```

**Check Poisson model**

```
## [1] 90.85094
```

```r
pearson.stat_pqi08 <- sum((na.omit(pqi08_data)$pqi08 - fitted(mod.pois_pqi08))^2/fitted(mod.pois_pqi08)
pearson.stat_pqi08/mod.pois_pqi08$df.residual # 90.84562
```

```
## [1] 90.84562
```

```r
# Evidence of overdispersion
mean(pqi08_data$pqi08, na.rm = T) #  1447.86
```

```
## [1] 1447.86
```

```r
var(pqi08_data$pqi08, na.rm = T) # 606848.1
```

```
## [1] 606848.1
```

Due to the evidence of overdispersion, we can say the negative binomial is the right choice.

```r
exp(coef(mod.nb_pqi08_1))
```

**Interpretation**

```
##                    (Intercept)             expansion_status2
##                    3.337991638                    1.160691240
##              expansion_status3             expansion_status4
##                    1.205621783                    1.173990405
##                   percent_male      percent_african_american
##                    0.004855028                    4.753737917
##               percent_hispanic                        age65-74
##                    1.472500611                    0.635477634
##                          age75+   percent_eligible_for_medicaid
##                    1.899931358                   26.060654335
## percent_eligible_for_medicaid_2
##                    0.007109440
```

```r
exp(confint(mod.nb_pqi08_1))
```

```
## Waiting for profiling to be done...
```

```
##                                       2.5 %      97.5 %
## (Intercept)                   1.7487279408   6.37112217
## expansion_status2             1.0958158961   1.22906501
## expansion_status3             1.1316335553   1.28434480
## expansion_status4             1.1093966675   1.24195864
## percent_male                  0.0011715313   0.02015097
## percent_african_american      3.6965575647   6.12163898
## percent_hispanic              1.1550949235   1.88157192
## age65-74                      0.6093185786   0.66275913
## age75+                        1.8218071945   1.98140390
## percent_eligible_for_medicaid 7.6493089640  88.13627171
```

```
## percent_eligible_for_medicaid_2 0.0005292642   0.09713554
```

Create plots for report

```
pqi05_data %>%
  na.omit() %>%
  ggplot(aes(x = expansion_status, y = pqi05, fill = expansion_status)) +
  geom_violin(alpha = 0.5, draw_quantiles = c(0.5)) +
  xlab("Expansion Status") +
  ylab("COPD or asthma admissions in older\n adults per 1,000 Beneficiaries") +
  scale_y_continuous(labels = scales::comma) +
  ggtitle("2018 Average PQI 05")
```
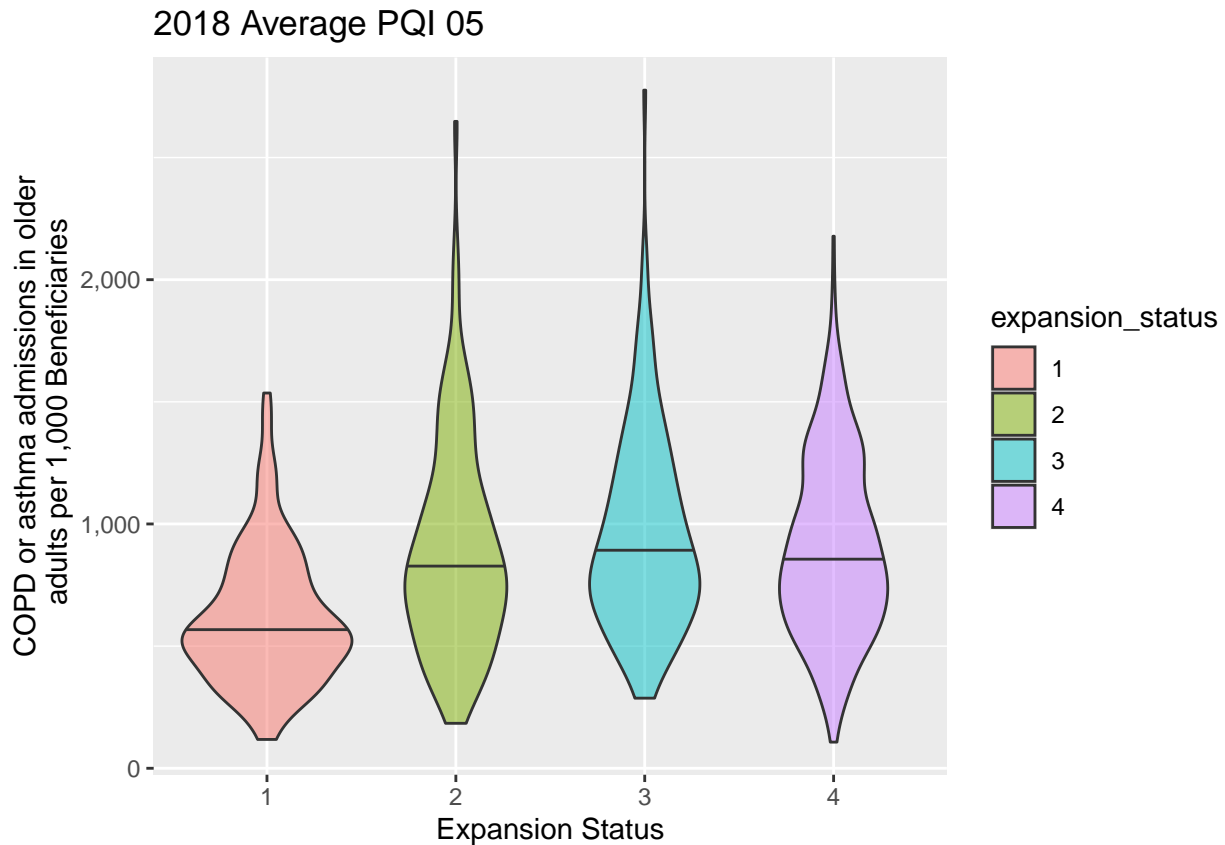


```
pqi08_data %>%
  na.omit() %>%
  ggplot(aes(x = expansion_status, y = pqi08, fill = expansion_status)) +
  geom_violin(alpha = 0.5, draw_quantiles = c(0.5)) +
  xlab("Expansion Status") +
  ylab("Congestive heart failure admissions per 1,000 beneficiaries") +
  scale_y_continuous(labels = scales::comma) +
  ggtitle("2018 Average PQI 08")
```

2018 Average PQI 08