Corey Huang
STAT 426 Project 1
4/15/2022

Ketorolac Report - STAT 426

Introduction:

In this study, I have been given the 'ketorolac.csv' dataset with the task of determining the effect that Toradol has on patient bed time and pain levels after surgery. In order to completely understand this study, we must first acquire an understanding of the clinical uses of Toradol. So, what is Toradol, or Ketorolac? Essentially, it is a painkiller used before and after surgical procedures. Unlike morphine, which is considered an opiate, Toradol is a nonsteroidal anti-inflammatory drug. According to professionals, the recommended span of Toradol use should be less than six days due to a dramatic increase for the risk of kidney damage after five days. From this, two main objectives are considered for this report. The first objective is to determine if taking Toradol decreases the patient's likelihood to stay in the hospital for more than five days. The second is to determine if taking Toradol decreases the patient's probability of having bad pain, rated on a scale from 1 to 10. Other variables in this dataset involve different statistics on the patient such as age, stage of cancer, and diagnosis, which provide additional information in regards to time of stay and pain levels. Please find in this report my findings and analysis of different variables and how Toradol ultimately affects patient stay time and pain levels.

The Data:

After loading the data, I created two scatterplots to visualize the distribution of the TOR predictor against the response LongStay and BadPain. Since the scatterplots created simple, four point graphs, it is clear that all three predictors that are being analyzed are categorical and binomial (Figure 1 & 2). Additionally, I created histogram plots of the frequency of LongStay and BadPain to determine the distribution of responses within each response. From the histograms, we can see that the distribution of both LongStay and BadPain is left-skewed (Figure 3 & 4). Next, I search for clear outliers in the dataset. Since every variable is categorical, I look for outliers in the continuous variable Morphine and find three observations that are outliers, which I drop from the dataset (Figure 5). Next, in order to determine confounding relationships within the predictor variables, I fit multiple generalized linear models with TOR being the response and every other variable being the predictor, individually. After fitting every model, I compute the confidence intervals of the odds-ratios for each model and look for the intervals that do not consist of 1. If there is no confounding interaction between two variables, then the odds of both will be close to a 1:1 ratio. Eventually, I found that the predictors that are confounding

variables against TOR are DOS.yr, Age, AddPro, OV, and Morphine (Figure 12). This means that use of Toradol differs based on the surgery year, age of patients, additional surgeries, ovary procedures, and morphine levels. This makes practical sense as well, as Morphine may act as a substitute to Toradol and different ages may have different pain tolerances. Additionally, different types of surgeries (OV) and the frequency of surgeries (AddPro) also make sense in affecting Toradol use.

Method:

I began the model fitting procedure by fitting a full generalized linear model with the logit link function and with LongStay and BadPain as the response and all other variables as predictors (Figure 6). I noticed that many of the predictors in the full model were not significant, so I decided to use a step function to do variable selection of only the significant models. Using backwards, forwards, and stepwise elimination, I find that the method that results in the smallest AIC is both backward and stepwise elimination, which both have an AIC of 85.08 (Figure 9, 10, and 11). After conducting backwards elimination, the reduced model becomes the model with General.diet, TOR, Comps, and Morphine as significant predictors to LongStay and BadPain. With the significant predictors known, I then fit three generalized linear models using the logit, probit, and cloglog link functions. In order to find which link function is optimal for this dataset, I plot the ROC curve of both LongStay and BadPain separately against the fitted values of the model. I do this six times in total, two of each predictor for three different link function models (Figure 15-20). After obtaining plots, I calculated the area under the curve of each ROC curve and averaged that area for every link function. Eventually, I conclude that the cloglog link function is the model with the greatest area under the curve, signifying that this is the link model that best describes the relationship in the dataset (Figure 21). Once the optimal model is obtained, I compute model diagnostics to find any highly influential points or outliers using Cook's Distance. From this, I find that there are two observations that can be classified as outliers using Cook's Distance (Figure 13). Additionally, I find that there are no highly influential points exceeding an absolute value of one (Figure 14). After discovering these outliers, I decided to drop them from my dataset and reuse the same optimal cloglog model with the new dataset. Finally, I found the overdispersion ratio to be 0.6740945, which is less than 1 and can therefore conclude that overdispersion is not present in the model.

Results:

The results of the final regression model are displayed below:

```
##
## Call:
## glm(formula = cbind(LongStay, BadPain) ~ General.diet + TOR +
##     Comps + Morphine, family = binomial(link = ("cloglog")),
##     data = keto2)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.051   0.000   0.000   0.227   1.334
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.853085   0.663846  -1.285  0.19877
## General.diet  0.400687   0.134867   2.971  0.00297 **
## TOR          -0.414385   0.328089  -1.263  0.20658
## Comps         0.923004   0.502723   1.836  0.06636 .

 ## Morphine     -0.005292   0.003026  -1.749  0.08030 .
 ## ---
 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 ##
 ## (Dispersion parameter for binomial family taken to be 1)
 ##
 ##     Null deviance: 65.458  on 69  degrees of freedom
 ## Residual deviance: 43.816  on 65  degrees of freedom
 ## AIC: 78.769
 ##
 ## Number of Fisher Scoring iterations: 6
```

```
##  (Intercept) General.diet          TOR        Comps     Morphine
##    0.4260982    1.4928505    0.6607466    2.5168387    0.9947223
```

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

```
##                   2.5 %   97.5 %
## (Intercept)  0.1086241 1.447033

## General.diet 1.1766904 1.932508
## TOR          0.3368932 1.259611
## Comps        0.9416194 7.619045
## Morphine     0.9887087 1.000370
```

From the outputs, we can conclude from the odds-ratios that Toradol and Morphine are the only predictors in the model that result in a decrease in LongStay and BadPain. As TOR increases, LongStay and BadPain are expected to increase by a factor of 0.6607466, which ultimately results in a decrease in the responses. Additionally, Morphine almost has no effect on the response since it's at an odds-ratio of 0.9947223, and General.diet and Comps significantly increase LongStay and BadPain with odds-ratios of 1.4928505 and 2.5168387.

Summary:

To summarize this study, the ‘ketorolac.csv’ dataset was given to me with the task of determining the effect that Toradol has on patient bed time and pain levels after surgery. According to professionals, the span of Toradol use is recommended to be less than six days because of a dramatic increase for the risk of kidney damage after five days. Through multiple linear regression procedures, I was able to find the optimal model in determining what affects patients to have a long stay time and bad pain levels. I found that the number of days until a patient's normal diet is consumed, the use of Toradol, surgical complications, and the amount of morphine administered to the patients to be the significant predictors that affect long stay time and bad pain levels. This led me to conclude that the use of Toradol decreases a patient’s chance of having to stay longer than five days and also decreases a patient's chance of having a pain level that’s higher than four. Interestingly, Morphine had less of a chance to decrease a patient’s stay time and pain levels, which may indicate that Toradol has more efficient usage in hospitals for both room availability and patient pain levels. Please find this report of my findings and analysis of different variables and how Toradol ultimately affects patient stay time and pain levels to be informative and helpful.

Appendix:

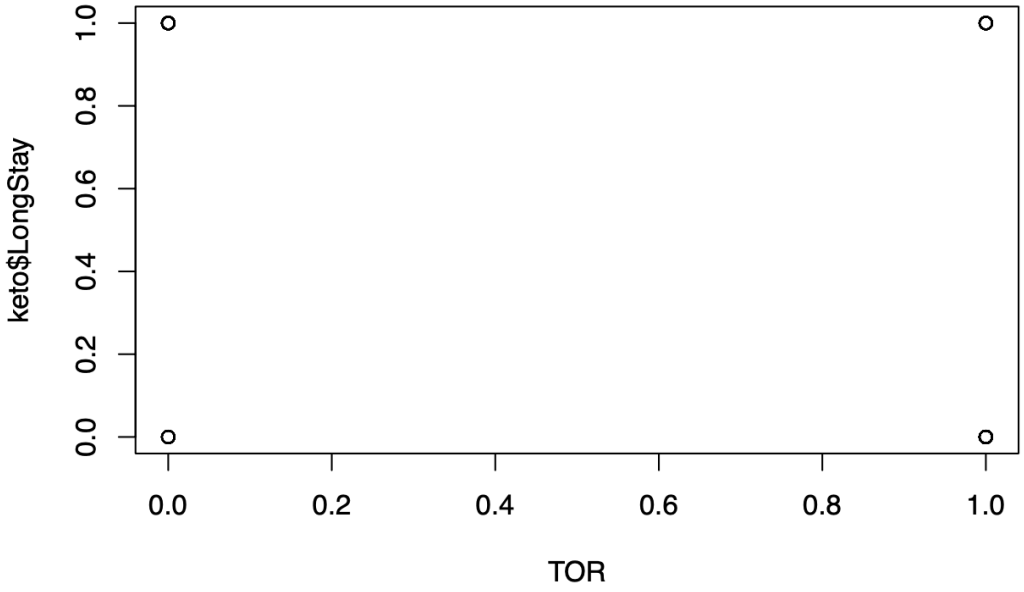

Figure 1: Toradol vs. LongStay

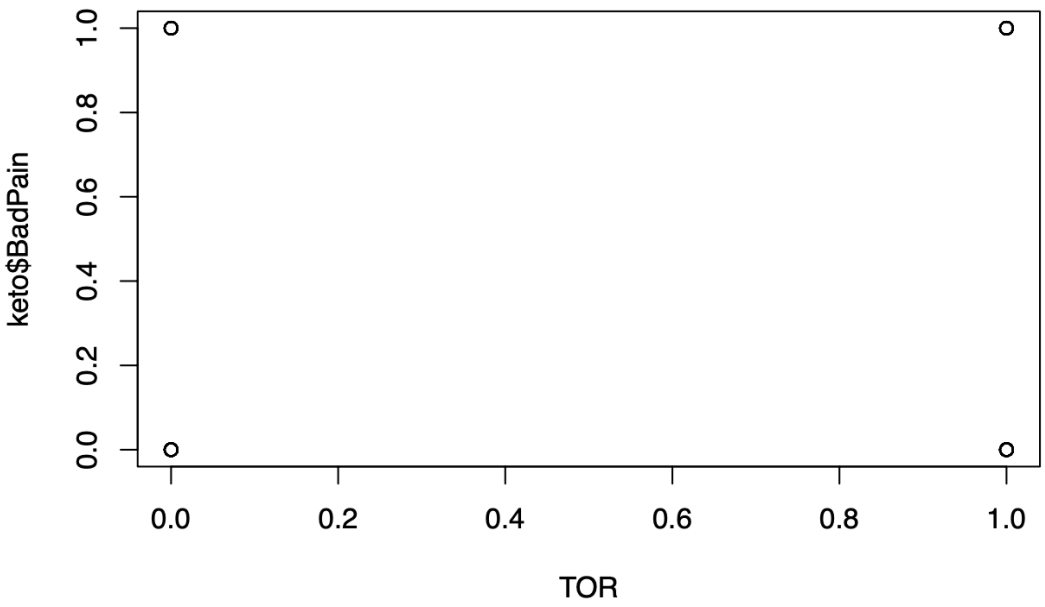

Figure 2: Toradol vs. BadPain

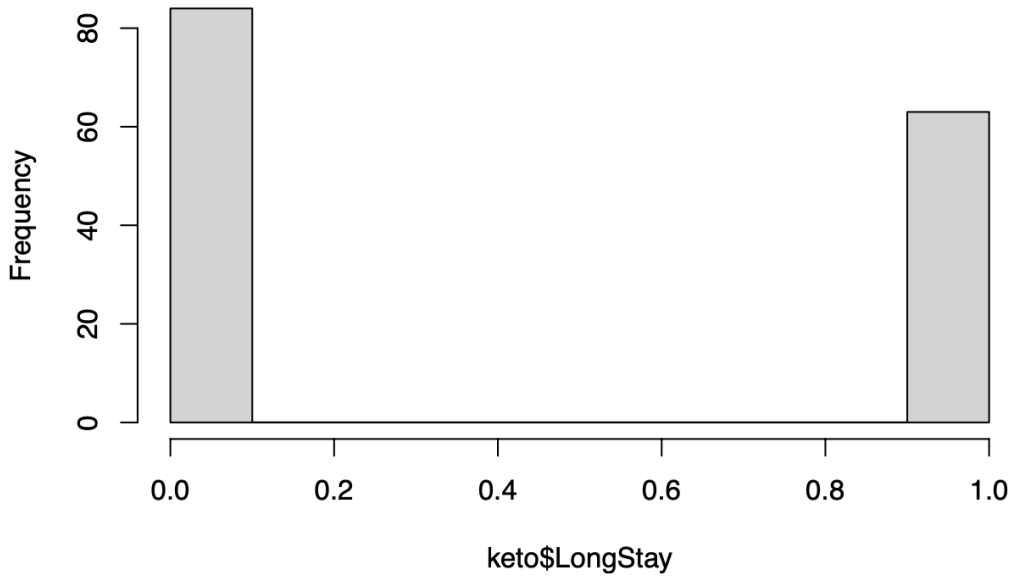

Figure 3: Histogram of LongStay

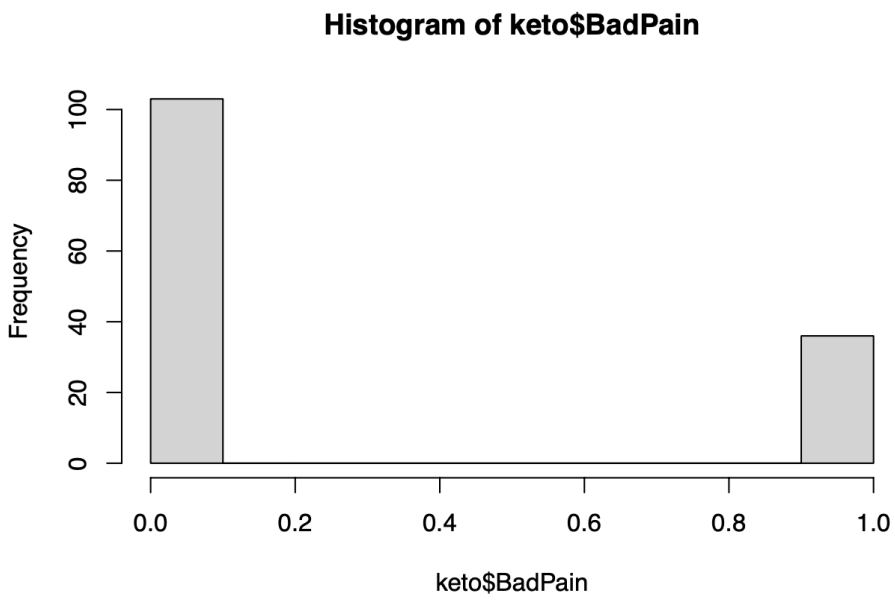

Figure 4: Histogram of BadPain

```
## [1] 317.5 310.3 300.4
```

Figure 5: Outliers in Morphine

```
##
## Call:
## glm(formula = cbind(LongStay, BadPain) ~ ., family = binomial,
##     data = keto1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7255   0.0000   0.0000   0.2567   1.2987
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   54.041627 104.784382   0.516   0.6060

## Pt.No         -0.006780   0.019016  -0.357   0.7214
## DOS.yr        -0.578857   1.094731  -0.529   0.5970
## Age           -0.001926   0.031971  -0.060   0.9520
## duration       0.095058   0.437566   0.217   0.8280
## Diagnosis      0.759088   0.954004   0.796   0.4262
## Stage         -0.037662   0.120613  -0.312   0.7548
## General.diet   0.759080   0.313642   2.420   0.0155 *
## HYS            0.258095   1.037276   0.249   0.8035
## OV             0.022926   0.429407   0.053   0.9574
## LNS            0.049338   0.427512   0.115   0.9081
## TOR           -0.171332   0.750212  -0.228   0.8194
## AddPro         0.004029   0.661664   0.006   0.9951
## Comps          1.887480   1.123521   1.680   0.0930 .
## Morphine      -0.014876   0.007671  -1.939   0.0525 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 67.597  on 71  degrees of freedom
## Residual deviance: 40.403  on 57  degrees of freedom
## AIC: 96.742
##
## Number of Fisher Scoring iterations: 5
```

Figure 6: Full Model Summary

```
##  (Intercept)        Pt.No       DOS.yr          Age     duration    Diagnosis
## 2.951076e+23 9.932429e-01 5.605385e-01 9.980760e-01 1.099722e+00 2.136326e+00
##        Stage General.diet          HYS           OV          LNS          TOR
## 9.630386e-01 2.136310e+00 1.294462e+00 1.023191e+00 1.050575e+00 8.425415e-01
##       AddPro        Comps     Morphine
## 1.004037e+00 6.602706e+00 9.852339e-01
```

Figure 7: Odds-Ratio for predictors in Full Model

```
##                      2.5 %        97.5 %
## (Intercept)  5.836895e-66 2.567985e+115
## Pt.No        9.559623e-01  1.031016e+00
## DOS.yr       6.131213e-02  4.729852e+00
## Age          9.374303e-01  1.064077e+00
## duration     4.720558e-01  2.692958e+00
## Diagnosis    3.553400e-01  1.592384e+01
## Stage        7.612202e-01  1.230589e+00
## General.diet 1.207115e+00  4.206408e+00
## HYS          1.662139e-01  1.031557e+01
## OV           4.393084e-01  2.416860e+00
## LNS          4.430011e-01  2.419306e+00
## TOR          1.972641e-01  3.860608e+00
## AddPro       2.701885e-01  3.725181e+00
## Comps        8.993086e-01  8.004004e+01
## Morphine     9.692778e-01  9.994216e-01
```

Figure 8: Confidence Intervals for Odds-Ratio

```
## Call:  glm(formula = cbind(LongStay, BadPain) ~ General.diet + TOR +
##     Comps + Morphine, family = binomial, data = keto)
##
## Coefficients:
##  (Intercept)  General.diet           TOR         Comps      Morphine
##    -1.441857      0.790459     -0.754884      1.387654     -0.006159
##
## Degrees of Freedom: 74 Total (i.e. Null);  70 Residual
##   (9 observations deleted due to missingness)
## Null Deviance:       68.8
## Residual Deviance: 45.97     AIC: 85.08
```

Figure 9: Backwards Elimination

```
## Call:  glm(formula = cbind(LongStay, BadPain) ~ Pt.No + DOS.yr + Age +
##     duration + Diagnosis + Stage + General.diet + HYS + OV +
##     LNS + TOR + AddPro + Comps + Morphine, family = binomial,
##     data = keto)
##
## Coefficients:
##  (Intercept)         Pt.No        DOS.yr           Age      duration
##   -20.344140     -0.011117      0.193929      0.002126      0.082433
##    Diagnosis         Stage  General.diet           HYS            OV
##     0.500917     -0.058739      0.716419      0.103481      0.053900
##          LNS           TOR        AddPro         Comps      Morphine
##    -0.089733     -0.463211     -0.068752      1.490770     -0.006185
##
## Degrees of Freedom: 74 Total (i.e. Null);  60 Residual
##   (9 observations deleted due to missingness)
## Null Deviance:       68.8
## Residual Deviance: 44.35     AIC: 103.5
```

Figure 10: Forwards Elimination

```
##
## Call:  glm(formula = cbind(LongStay, BadPain) ~ General.diet + TOR +
##     Comps + Morphine, family = binomial, data = keto)
##
## Coefficients:
##  (Intercept)  General.diet           TOR         Comps      Morphine
##    -1.441857      0.790459     -0.754884      1.387654     -0.006159
##

  ## Degrees of Freedom: 74 Total (i.e. Null);  70 Residual
  ##   (9 observations deleted due to missingness)
  ## Null Deviance:       68.8
  ## Residual Deviance: 45.97     AIC: 85.08
```

Figure 11: Stepwise Elimination

```r
model.yr <- glm(TOR~DOS.yr, family="binomial", data=keto)
exp(cbind(coef(model.yr), confint(model.yr, level = 0.95)))
```

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

```
##                                 2.5 %       97.5 %
## (Intercept) 3.551549e-78 1.006172e-105 4.613098e-54
## DOS.yr      6.330440e+00  3.566386e+00 1.219626e+01
```

```r
model.stage <- glm(TOR~Stage, family="binomial", data=keto)
exp(cbind(coef(model.stage), confint(model.stage, level = 0.95)))
```

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

```
##                            2.5 %   97.5 %
## (Intercept) 1.6547233 1.0015080 2.762029
## Stage       0.9477105 0.8325012 1.078556
```

```r
model.diagnosis <- glm(TOR~Diagnosis, family="binomial", data=keto)
exp(cbind(coef(model.diagnosis), confint(model.diagnosis, level = 0.95)))
```

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

```
##                           2.5 %    97.5 %
## (Intercept) 3.396658 1.1677672 10.477395
## Diagnosis   0.605916 0.3315804  1.081428
```

```r
model.duration <- glm(TOR~duration, family="binomial", data=keto)
exp(cbind(coef(model.duration), confint(model.duration, level = 0.95)))
```

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

```
##                           2.5 %   97.5 %
## (Intercept) 1.211784 0.6131800 2.412387
## duration    1.076784 0.8038866 1.449365
```

```r
model.age <- glm(TOR~Age, family="binomial", data=keto)
exp(cbind(coef(model.age), confint(model.age, level = 0.95)))
```

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

```
##                            2.5 %    97.5 %
## (Intercept) 9.5801666 2.7603655 36.533319
## Age         0.9666964 0.9452775  0.987297
```

```r
model.addpro <- glm(TOR~AddPro, family="binomial", data=keto)
exp(cbind(coef(model.addpro), confint(model.addpro, level = 0.95)))
```

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

```r
model.HYS <- glm(TOR~HYS, family="binomial", data=keto)
exp(cbind(coef(model.HYS), confint(model.HYS, level = 0.95)))
```

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

```
##                            2.5 %   97.5 %
## (Intercept) 1.279070 0.8599230 1.914988
## HYS         1.346465 0.6698396 2.757111
```

```
model.OV <- glm(TOR~OV, family="binomial", data=keto)
exp(cbind(coef(model.OV), confint(model.OV, level = 0.95)))
```

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

```
##                              2.5 %     97.5 %
## (Intercept) 5.1701044 1.5525762 20.5582228
## OV          0.6094545 0.3732103  0.9458596
```

```
model.LNS <- glm(TOR~LNS, family="binomial", data=keto)
exp(cbind(coef(model.LNS), confint(model.LNS, level = 0.95)))
```

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

```
##                            2.5 %   97.5 %
## (Intercept) 1.332180 0.4160182 4.381105
## LNS         1.024231 0.6342673 1.646363
```

```
model.Morphine <- glm(TOR~Morphine, family="binomial", data=keto)
exp(cbind(coef(model.Morphine), confint(model.Morphine, level = 0.95)))
```

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

```
##                              2.5 %    97.5 %
## (Intercept) 3.2054638 1.7220434 6.2481848
## Morphine    0.9910085 0.9848184 0.9967179
```

```
model.Comps <- glm(TOR~Comps, family="binomial", data=keto)
exp(cbind(coef(model.Comps), confint(model.Comps, level = 0.95)))
```

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

```
##                              2.5 %   97.5 %
## (Intercept) 1.4814815 1.0521659 2.102182
## Comps       0.5785714 0.1773607 1.832716
```

Figure 12: Confidence Intervals for Odds-Ratio of predictors against TOR

```
##         31         56
## 0.0707130 0.4053234
```

Figure 13: Cook's Distances

```
##     (Intercept) General.diet   TOR Comps Morphine
## 2         FALSE        FALSE FALSE FALSE    FALSE
## 4         FALSE        FALSE FALSE FALSE    FALSE
## 5         FALSE        FALSE FALSE FALSE    FALSE
## 6         FALSE        FALSE FALSE FALSE    FALSE
## 7         FALSE        FALSE FALSE FALSE    FALSE
## 8         FALSE        FALSE FALSE FALSE    FALSE
## 9         FALSE        FALSE FALSE FALSE    FALSE
## 10        FALSE        FALSE FALSE FALSE    FALSE
## 11        FALSE        FALSE FALSE FALSE    FALSE
## 12        FALSE        FALSE FALSE FALSE    FALSE
## 13        FALSE        FALSE FALSE FALSE    FALSE
## 14        FALSE        FALSE FALSE FALSE    FALSE
## 15        FALSE        FALSE FALSE FALSE    FALSE
## 17        FALSE        FALSE FALSE FALSE    FALSE
## 18        FALSE        FALSE FALSE FALSE    FALSE
## 19        FALSE        FALSE FALSE FALSE    FALSE
## 21        FALSE        FALSE FALSE FALSE    FALSE
## 22        FALSE        FALSE FALSE FALSE    FALSE
## 24        FALSE        FALSE FALSE FALSE    FALSE
## 25        FALSE        FALSE FALSE FALSE    FALSE
## 27        FALSE        FALSE FALSE FALSE    FALSE
## 28        FALSE        FALSE FALSE FALSE    FALSE
## 29        FALSE        FALSE FALSE FALSE    FALSE
## 30        FALSE        FALSE FALSE FALSE    FALSE
## 31        FALSE        FALSE FALSE FALSE    FALSE
## 32        FALSE        FALSE FALSE FALSE    FALSE
## 33        FALSE        FALSE FALSE FALSE    FALSE
## 37        FALSE        FALSE FALSE FALSE    FALSE
## 38        FALSE        FALSE FALSE FALSE    FALSE
## 39        FALSE        FALSE FALSE FALSE    FALSE
## 41        FALSE        FALSE FALSE FALSE    FALSE
## 43        FALSE        FALSE FALSE FALSE    FALSE
## 44        FALSE        FALSE FALSE FALSE    FALSE
## 45        FALSE        FALSE FALSE FALSE    FALSE
## 46        FALSE        FALSE FALSE FALSE    FALSE
## 47        FALSE        FALSE FALSE FALSE    FALSE
## 49        FALSE        FALSE FALSE FALSE    FALSE
## 52        FALSE        FALSE FALSE FALSE    FALSE
## 53        FALSE        FALSE FALSE FALSE    FALSE
## 54        FALSE        FALSE FALSE FALSE    FALSE
## 55        FALSE        FALSE FALSE FALSE    FALSE
```

Figure 14: Stepwise Elimination

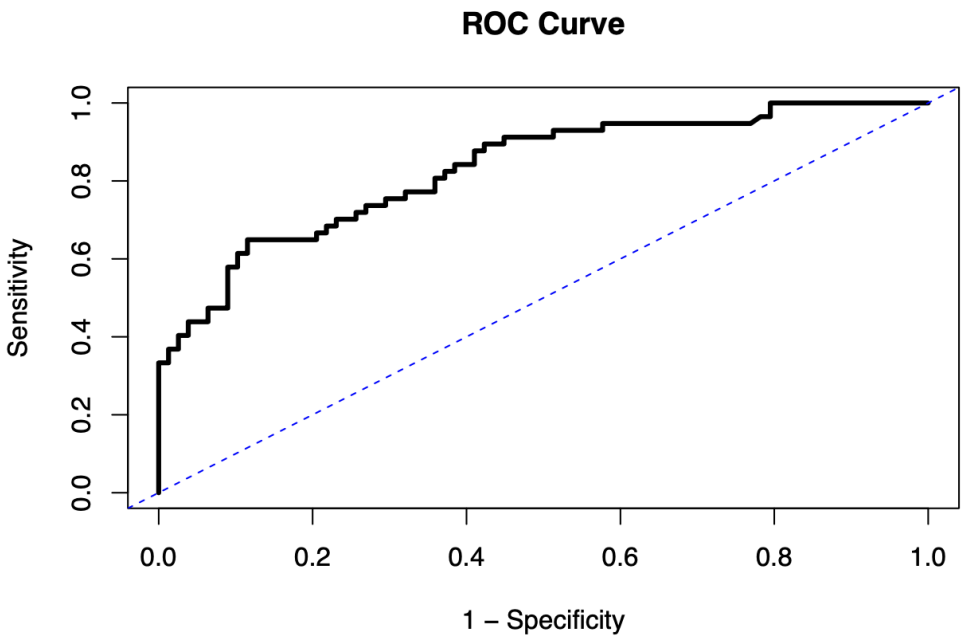

Figure 15: ROC Curve for LongStay on Logit Model

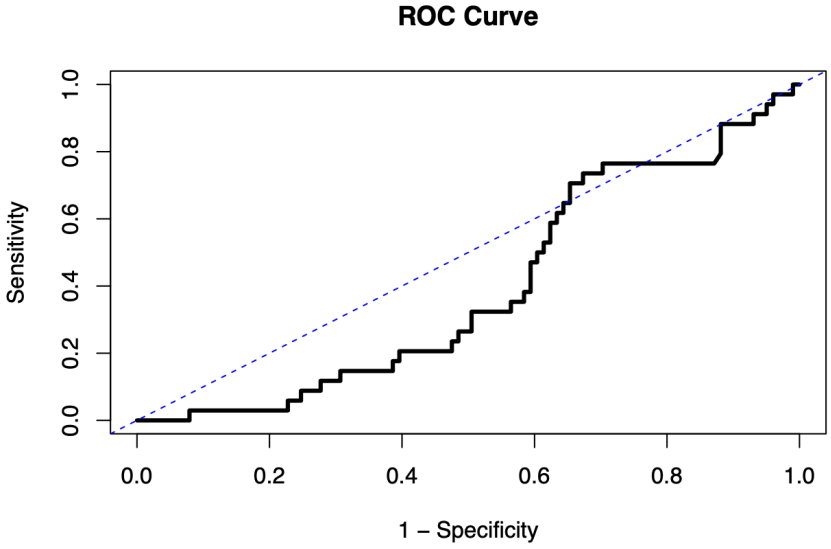

Figure 16: ROC Curve for BadPain on Logit Model

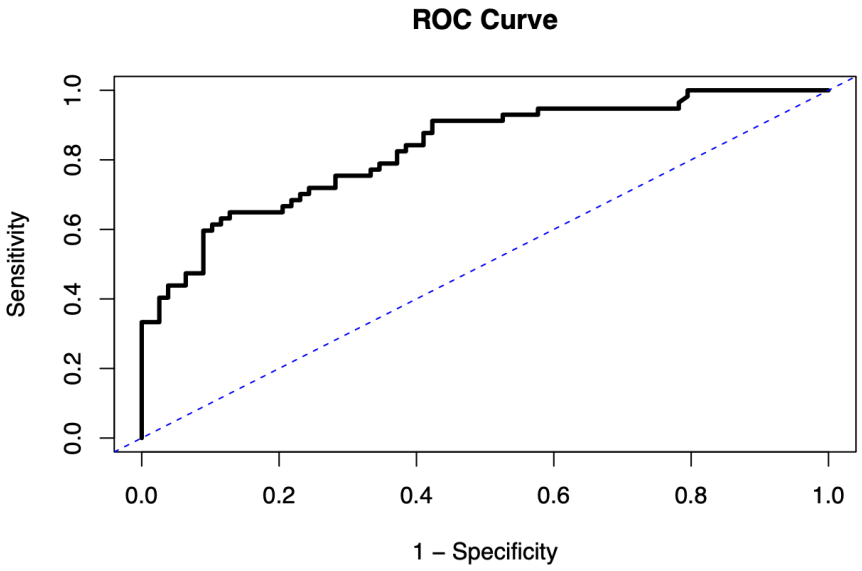

Figure 17: ROC Curve for LongStay on Probit Model

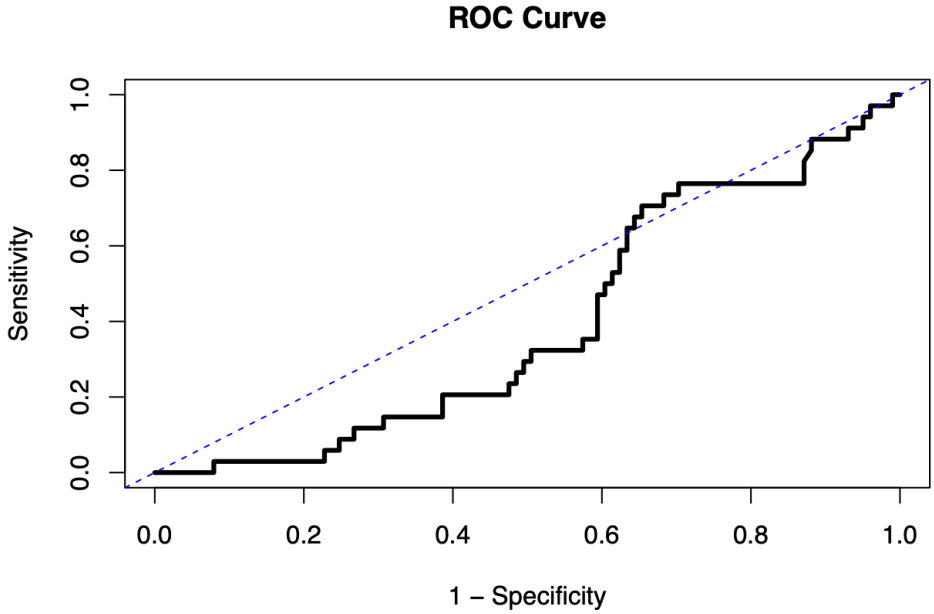

Figure 18: ROC Curve for BadPain on Probit Model

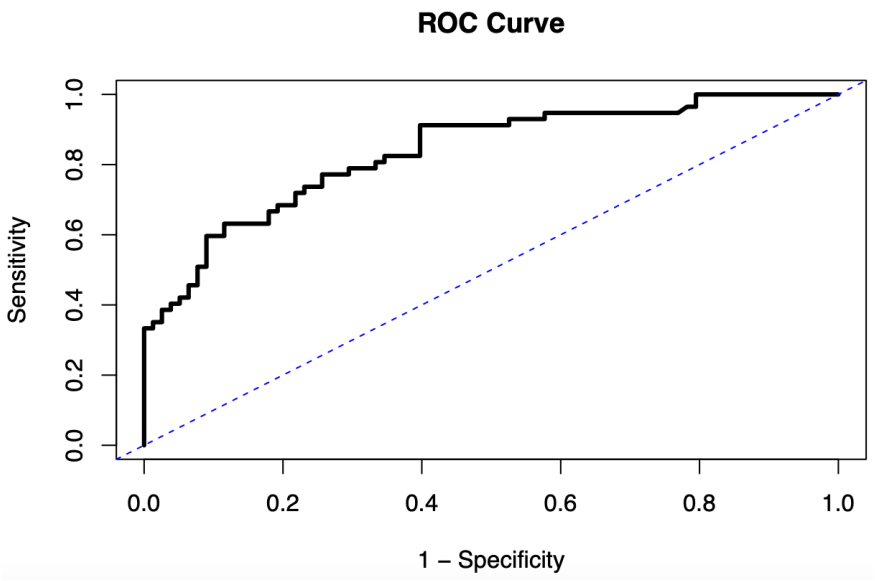

Figure 19: ROC Curve for LongStay on Cloglog Model

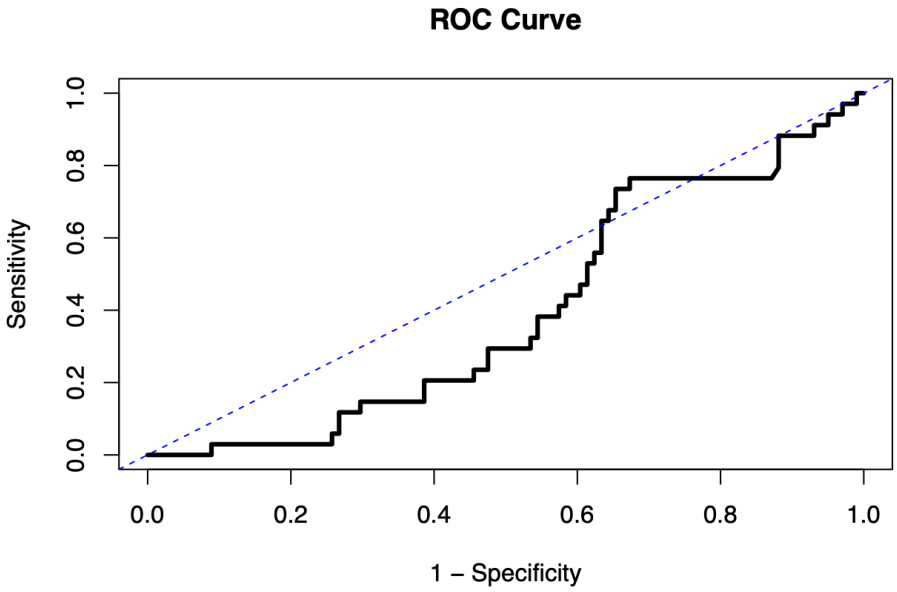

Figure 20: ROC Curve for BadPain on Cloglog Model

Logit: 0.61222185 Probit: 0.6126918 Cloglog: 0.61678055

Figure 21: Average Area under ROC Curve for Each Link Function

```
##  (Intercept) General.diet         TOR       Comps    Morphine
##    0.4260982    1.4928505   0.6607466   2.5168387   0.9947223
```

Figure 22: Odds-Ratio of New Reduced Model

```
## Waiting for profiling to be done
##                   2.5 %   97.5 %
## (Intercept)  0.1086241 1.447033
## General.diet 1.1766904 1.932508
## TOR          0.3368932 1.259611
## Comps        0.9416194 7.619045
## Morphine     0.9887087 1.000370
```

Figure 23: Confidence Interval of Odds-Ratios

```
## [1] 0.6740945
```

Figure 24: Overdispersion Ratio

```
##
## Call:
## glm(formula = cbind(LongStay, BadPain) ~ General.diet + TOR +
##     Comps + Morphine, family = binomial(link = ("cloglog")),
##     data = keto2)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.051   0.000   0.000   0.227   1.334
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.853085   0.663846  -1.285  0.19877
## General.diet  0.400687   0.134867   2.971  0.00297 **
## TOR          -0.414385   0.328089  -1.263  0.20658
## Comps         0.923004   0.502723   1.836  0.06636 .

## Morphine     -0.005292   0.003026  -1.749  0.08030 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 65.458  on 69  degrees of freedom
## Residual deviance: 43.816  on 65  degrees of freedom
## AIC: 78.769
##
## Number of Fisher Scoring iterations: 6
```

Figure 25: Final Model Summary