An Analysis of the Risk Factors Associated with Glaucomatous Progression

ABSTRACT:

Glaucoma is one of the premier causes of blindness in the United States. Glaucoma is a group of conditions that irreversibly damage the optic nerve over time. Regular eye exams are imperative for glaucoma treatment, as its slow progression can lead the affected into advanced stages of vision loss before it is noticed. However, once diagnosed, glaucomatous progression can be slowed and vision loss can be prevented. Optometrists use a variety of tests to diagnosis, including tonometry (the procedure of determining intraocular pressure), pachymetry (measuring cornea thickness), and perimetry (a visual field test). I will fit a multiple logistic regression model to the data to determine the significant predictors of glaucomatous progression. In this analysis, I find that those in the treatment group are 3.25 times less likely to have an indication of progressing glaucoma than those in the control group, and that low blood pressure has a higher association with progressing glaucoma than high blood pressure.

INTRODUCTION:

In this project, the data is a subset of observations from a study of treatment on a glaucomatous progression outcome. Each of these 1639 subjects presented one randomly selected eye in this data set, and had a number of anatomical, physiological, and clinical parameters recorded.

In this paper, I will explore the following questions: (1) Is the treatment effective? (2) Is high blood pressure a bigger risk for glaucomatous progression than low blood pressure? (3) Which type of vision test is most associated with glaucomatous progression?

I'll hypothesize that the treatment is indeed effective in stopping glaucomatous progression. Also I will hypothesize that high blood pressure is a bigger risk for glaucoma progression than low blood pressure. Additionally, I expect that Cup-to-Disk Ratio and the Vertical Cup-to-Disk Ratio will carry different levels of significance, as well as pattern standard deviation and corrected pattern standard deviation. My last hypothesis is that determining intraocular pressure through tonometry is the best predictor of glaucoma progression, because high intraocular pressure is known to be a risk factor for glaucoma.

Determining the significance of the independent variables on progression of glaucoma will require us to use of regression analysis. As the response variable is a dichotomous indicator variable, I will create a multinomial logistic regression model to determine the significance that each variable has on glaucomatous progression.

SUMMARY INFO:

Below are the contingency tables along with odds ratios and confidence intervals, this is also in **APPENDIX-1**.

	PGON		Odds Ratio			PGON		Odds Ratio	
Eye	0	1	**OR = 1.2493**		**Gender**	0	1	**OR = 1.83**	
0	774	64	Lower = .8817		0	883	61	Lower = 1.289	
1	726	75	Upper = 1.7704		1	617	78	Upper = 2.5979	
	PGON		Odds Ratio			PGON		Odds Ratio	
TRT	0	1	**OR = .3532**		**History**	0	1	**OR = 1.1499**	
0	740	102	Lower = .2392		0	987	87	Lower = .8025	
1	160	37	Upper = .5214		1	513	52	Upper = 1.6478	
	PGON		Odds Ratio			PGON		Odds Ratio	
Race	0	1	**OR = 1.2416**		**Sysbb**	0	1	**OR = .699**	
0	1134	97	Lower = .9168		0	1439	135	Lower = .2503	
1	366	42	Upper = 1.9632		1	61	4	Upper = 1.9517	
	PGON		Odds Ratio			PGON		Odds Ratio	
Ccb	0	1	**OR = 1.2808**		**Heart**	0	1	**OR = 2.0392**	
0	1326	119	Lower = .7774		0	1416	124	Lower = 1.1426	
1	174	20	Upper = 2.1102		1	84	15	Upper = 3.6392	
	PGON		Odds Ratio			PGON		Odds Ratio	
MIGR	0	1	**OR = 1.0744**		**HPB**	0	1	**OR= 1.2895**	
0	1338	123	Lower = .6225		0	944	79	Lower = .9072	
1	162	16	Upper = 1.854		1	556	60	Upper = 1.833	
	PGON		Odds Ratio			PGON		Odds Ratio	
LPB	0	1	**OR = 1.7979**		**DIAB**	0	1	**OR = .3805**	

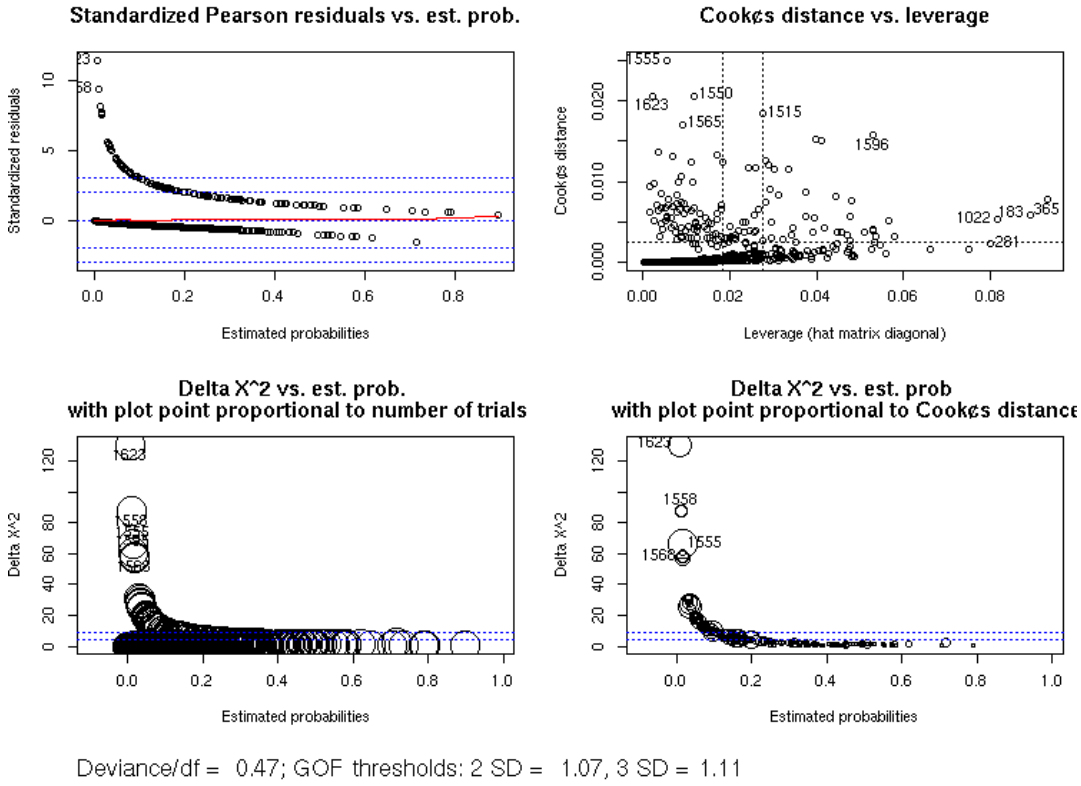
0	1438	129	Lower = .9002		0	1341	133	Lower = .1652
1	62	10	Upper = 3.591		1	159	6	Upper = .8764
	PGON		Odds Ratio			PGON		Odds Ratio
Stroke	0	1	**OR = .7174**		**Myopic**	0	1	**OR = 1.0464**
0	1485	138	Lower = .0941		0	976	89	Lower = .7284
1	15	1	Upper = 5.4718		1	524	50	Upper = 1.5033

Below are the summaries for the three different fitted models, along with p-values and regression coefficients, this is also in **APPENDIX-3**. The predicted odds ratios for the final model are shown as well.

	Univariate Model			Search Model			Best Fit Model			
Variable	Est.	Std.	p-val	Est.	Std.	p-val	Est.	St.d	p-val	OR-hat
Intercept	-4.502	2.092	0.031	-4.617	2.008	0.021	-4.962	2.020	0.014	0.007
Treatment	-1.175	0.214	4.00E-008	-1.176	0.214	3.78E-008	-1.177	0.214	3.82E-008	0.308
Race	-0.183	0.234	0.435							
Gender	0.534	0.203	0.008	0.522	0.200	0.009	0.547	0.201	7.00E-003	1.727
History	0.322	0.203	0.112	0.300	0.202	0.137	0.318	0.202	0.116	1.375
Low BP	1.068	0.406	0.009	0.985	0.401	0.014	1.083	0.407	0.008	2.952
Heart D	0.527	0.344	0.126	0.582	0.341	0.088	0.53	0.344	0.13	1.695
High BP	0.345	0.207	0.096				0.307	0.202	0.128	1.360
Diabetes	-1.090	0.452	0.016	-1.089	0.452	0.016	-1.121	0.450	0.013	0.326
Myopia	0.565	0.216	0.009	0.560	0.215	0.009	0.573	0.215	0.008	1.774
IOP	0.162	0.032	3.95E-007	0.160	0.032	4.29E-007	0.162	0.032	3.01E-007	1.176
CCT	-0.013	0.003	4.63E-006	-0.013	0.003	3.35E-006	-0.013	0.003	5.20E-006	0.988

CD	-0.453	0.990	0.647							
VCD	3.200	0.965	0.001	2.781	0.522	1.02E-007	2.779	0.524	1.16E-007	16.099
PSD	1.026	0.322	0.001	1.011	0.317	0.001	1.028	0.319	0.001	2.796
MD	-0.186	0.090	0.038	-0.184	0.089	0.039	-0.181	0.089	0.043	0.834
Age	0.031	0.011	0.006	0.033	0.011	0.003	0.032	0.011	0.003	1.033

Below are the diagnostic plots for my model, also found in **APPENDIX-7**



Deviance/df = 0.47; GOF thresholds: 2 SD = 1.07, 3 SD = 1.11

STATISTICAL ANALYSIS:

I started with a univariate analysis, that is, I compared each predictor variable to our response variable (progressive glaucomatous optic neuropathy indicator) to determine each variable's significance. I placed our categorical variables into contingency tables to calculate the odds ratio (OR) for each and corresponding

confidence intervals. The odds ratio is the ratio of an event occurring in one group to the odds of that event occurring in another group. The OR confidence interval of most significant variables will not include 1, however we will still consider including variables in our model if the lower bound of their OR confidence intervals are larger than 0.9. Our main variable of interest, treatment, has a very significant odds ratio of 0.3532.

Additionally, gender, heart, and diabetes have high odds ratios, indicating significance. I will also include race, high blood pressure, and low blood pressure in our model because their odds ratios have lower confidence bounds that are larger than 0.9. Though eye, family history, systemic beta-blocker use, calcium channel blocker, the migraine indicator, the stroke indicator, and the myopia indicator didn't make the confidence interval cut, I will still include the myopia indicator and family history of glaucoma indicator because they intuitively seem like significant predictors. The contingency tables, odds ratios, and corresponding confidence intervals for these variables can be found in **APPENDIX-1**.

There are different methods of univariate analysis that I can use to determine which of our continuous variables to include in our logistic regression model. The univariate Wald statistic and Welch's two-sample t-test produce nearly equivalent results at the univariate level, however for this analysis I will only use the results from the two-sample t-test, which hypothesizes that the variables cannot be used to predict PGON. A low p-value rejects this hypothesis, indicating that the variable is significant for our analysis.

The p-values that we obtain for Welch's two-sample t-test indicate that 6 of our 8 continuous covariates are very significant, while mean deviation and corrected pattern standard deviation aren't significant at the alpha = .0001 level. These p-values can be seen in **APPENDIX-2**. This answers our question about whether corrected pattern standard deviation is a better predictor than pattern standard deviation. Additionally, the vertical cup-to-disk ratio is much more significant than the cup-to-disk ratio. I will remove corrected pattern standard deviation from our model, but include mean deviation because this perimetry statistic may include some additional significance in our test model. Now with our 10 discrete variables and 6 continuous variables, we have our first "test" logistic regression model:

"Test Model" = PGON~logistic(Treatment, Race, Gender, History, Low Blood Pressure, Heart Disease, High Blood Pressure, Diabetes, Myopia, Intraocular Pressure, Central Corneal Thickness, Cup-to-Disk Ratio, Vertical Cup-to-Disk Ratio, Mean Deviation, Pattern Standard Deviation, Age), which has an AIC of 793.5681. The summary information and regression coefficients for this model are found in **APPENDIX-3.1**

The Akaike Information Criterion (AIC) can be used to find the best fit model from the variables in our test model. In the test model, our AIC was 793.5681. A lower AIC is associated with a better model fit. An exhaustive, time-consuming multiple general linear model search goes through every possible model to estimate the relative quality of each model compared to every other possible model in our search. We eventually obtain

the following model with the lowest AIC of 790.6808:

"Search Model" = PGON~logistic(Treatment, Gender, History, Low Blood Pressure, Heart Disease, Diabetes, Myopia, Intraocular Pressure, Central Corneal Thickness, Vertical Cup-to-Disk Ratio, Mean Deviation, Pattern Standard Deviation, Age). The summary information and regression coefficients for this model are found in **APPENDIX-3.2**.

Finally, I used backward elimination, which is a model-fitting technique in which one starts with the full model and eliminates insignificant variables one-by-one, until obtaining a model with only significant variables.

"Backward Elimination Model" = PGON~logistic(Treatment, Gender, History, Low Blood Pressure, Heart Disease, High Blood Pressure, Diabetes, Myopia, Intraocular Pressure, Central Corneal Thickness, Vertical Cup-to-Disk Ratio, Mean Deviation, Pattern Standard Deviation, Age).

Usually, the AIC search finds the model with the lowest AIC. In this case however, a lower AIC was found when backward elimination decided not to remove high blood pressure from the rest of the variables in our best search. The AIC for this new model is 790.3834, seen in **APPENDIX-6**. Because the model found using backward elimination has the lowest AIC, this is the best fitting model.

Now I checked to see if there were any significant interaction effects that should be included, and checked to see if any of the continuous covariates needed a transformation. Upon doing a search for interaction effects between the most significant variables in our model, it was determined that there are no significant interaction effects to include in this model. To determine if any of the six continuous variables in this model needed to be transformed, I plotted the data points against their standardized pearson residuals and determined that the residuals appear to be normally distributed, so no transformations will be necessary. These plots are found in **APPENDIX-4**.

Additionally, I used a scatterplot matrix to determine relationships between the continuous variables. As we can see in **APPENDIX-5**, pattern standard deviation and corrected pattern standard deviation have a linear relationship like we'd expected, as well as cup-to-disk ratio and vertical cup-to-disk ratio.

As there are no significant interaction effects and none of the continuous variables require a transformation, we can now proceed to look at the diagnostics of our model and check for outliers, leverage, and influence.

I"ll use four more plots for my model diagnostic analysis. The standardized pearson residuals vs estimated probability plot is used to ascertain whether we have any outliers in the model. The cook's distance vs. leverage plot detects highly influential points in the model. The change in pearson chi-square vs estimated

probabilities to check whether the model fits reasonably well. The proportionality of the points to number of trials and cook's distance allows one to more clearly determine the contributions of residual and leverage to number of trials and cook's distance, respectively. These diagnostic plots are found in **APPENDIX-6**.

Looking at the diagnostics, one can see that observations 1555 and 1623 have large cook's distances, large pearson residuals, and big changes in chi-squared. Additionally, observation 1558 has a large pearson residual and large changes in chi-squared; however, this is not enough information to consider removing these observations from our model, as we don't have any evidence that the data for these observations was properly or improperly collected.

The Hosmer-Lemeshow Test for logistic regression is a "goodness of fit" test, letting me know how well the selected model fits the data. The H-L test statistic approximately follows the chi-squared distribution with g – 2 degrees of freedom, when g > p +1, where g is the number of groups, and p is the number of predictors in the model. Therefore, I can calculate the p-value using the area under the curve of the chi-squared distribution to assess for goodness-of-fit. A small p-value indicates poor fit. The lowest p-value I obtained using different numbers of groups was 0.1073, when using 15 groups. Therefore, we can conclude that this model is not poorly fit at all standard levels of significance. This is shown in **APPENDIX-7.**

I also decided to use the Likelihood Ratio Test, which tests a reduced model against a full model that is saturated with every variable, as a way to determine if there is any additional significance in the variables that were omitted. Then, I tested against the chi-squared distribution and obtained a p-value of 0.929, again indicating that the model is a very good fit. The results from this test can be found in **APPENDIX-7**.

CONCLUSIONS:

With these results I can now answer the questions I asked in the beginning of this report. Because the predicted odds ratio for treatment is .308, that means that those in the treatment group are 1/.308 = 3.25 times less likely to have progressing glaucoma than those in the control group. Thus, we can conclude that this treatment is effective. The predicted odds ratio for low blood pressure in the final model is 2.952, which is significantly higher than the predicted odds ratio for high blood pressure at 1.36. Therefore, those with low blood pressure have a stronger association with glaucoma progression than those with high blood pressure. Lastly, I compare the odds ratios of the variables intraocular pressure, central corneal thickness, vertical cup-to-disk ratio, mean deviation, and pattern standard deviation to determine which of the vision examinations is most associated with progressive glaucomatous optic neuropathy. As pattern standard deviation has a large odd ratio of 2.796, I conclude that a field vision test called pelimetry is the most effective of the vision exams.

One of the main limitations of this study is that there is no given information on the nature of the treatment. With a better understanding of the dependent variable, I'd be able to do a more in-depth analysis and

have a better idea on which variables are significant and might interact.

In future studies, additional information on the treatment protocol will help achieve the most accurate results. The data could be expanded by including other vision examination statistics and a broader category for race than just an African-American indicator. Finally, in a future prospective study I would categorize glaucomatous progression with multiple stages to gather very useful data on the long-term progression of glaucoma.

APPENDIX

APPENDIX-1:

	PGON		Odds Ratio			PGON		Odds Ratio	
Eye	0	1	**OR = 1.2493**		**Gender**	0	1	**OR = 1.83**	
0	774	64	Lower = .8817		0	883	61	Lower = 1.289	
1	726	75	Upper = 1.7704		1	617	78	Upper = 2.5979	
	PGON		Odds Ratio			PGON		Odds Ratio	
TRT	0	1	**OR = .3532**		**History**	0	1	**OR = 1.1499**	
0	740	102	Lower = .2392		0	987	87	Lower = .8025	
1	160	37	Upper = .5214		1	513	52	Upper = 1.6478	
	PGON		Odds Ratio			PGON		Odds Ratio	
Race	0	1	**OR = 1.2416**		**Sysbb**	0	1	**OR = .699**	
0	113	97	Lower = .9168		0	143	135	Lower = .2503	

	4					9		
1	366	42	Upper = 1.9632		1	61	4	Upper = 1.9517
	PGON		Odds Ratio			PGON		Odds Ratio
Ccb	0	1	**OR = 1.2808**		**Heart**	0	1	**OR = 2.0392**
0	1326	119	Lower = .7774		0	1416	124	Lower = 1.1426
1	174	20	Upper = 2.1102		1	84	15	Upper = 3.6392
	PGON		Odds Ratio			PGON		Odds Ratio
MIGR	0	1	**OR = 1.0744**		**HPB**	0	1	**OR= 1.2895**
0	1338	123	Lower = .6225		0	944	79	Lower = .9072
1	162	16	Upper = 1.854		1	556	60	Upper = 1.833
	PGON		Odds Ratio			PGON		Odds Ratio
LPB	0	1	**OR = 1.7979**		**DIAB**	0	1	**OR = .3805**
0	1438	129	Lower = .9002		0	1341	133	Lower = .1652
1	62	10	Upper = 3.591		1	159	6	Upper = .8764
	PGON		Odds Ratio			PGON		Odds Ratio
Stroke	0	1	**OR = .7174**		**Myopic**	0	1	**OR = 1.0464**
0	1485	138	Lower = .0941		0	976	89	Lower = .7284
1	15	1	Upper = 5.4718		1	524	50	Upper = 1.5033

APPENDIX-2:

```
> t.test(IOP[PGON==0], IOP[PGON==1])

     Welch Two Sample t-test

data:  IOP[PGON == 0] and IOP[PGON == 1]
t = -5.1426, df = 159.51, p-value = 7.852e-07
alternative hypothesis: true difference in me
95 percent confidence interval:
 -2.0423589 -0.9089457
sample estimates:
mean of x mean of y
 25.04233  26.51799

> t.test(CCT[PGON==0], CCT[PGON==1])

     Welch Two Sample t-test

data:  CCT[PGON == 0] and CCT[PGON == 1]
t = 5.9877, df = 165.89, p-value = 1.279e-08
alternative hypothesis: true difference in me
95 percent confidence interval:
 13.56246 26.90672
sample estimates:
mean of x mean of y
 573.3065  553.0719

> t.test(CD[PGON==0], CD[PGON==1])

     Welch Two Sample t-test

data:  CD[PGON == 0] and CD[PGON == 1]
```

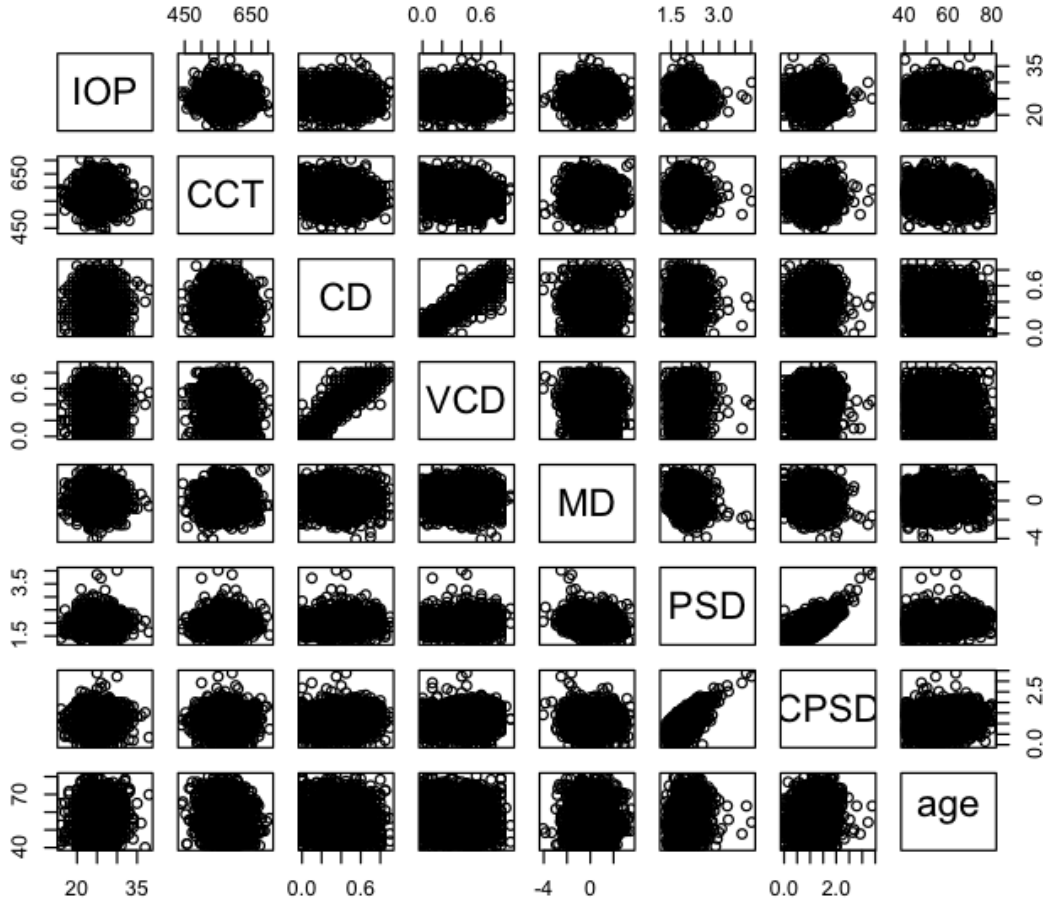
```
> t.test(MD[PGON==0], MD[PGON==1])

     Welch Two Sample t-test

data:  MD[PGON == 0] and MD[PGON == 1]
t = 3.3896, df = 167.65, p-value = 0.0008726
alternative hypothesis: true difference in me
95 percent confidence interval:
 0.1399159 0.5302286
sample estimates:
 mean of x  mean of y
 0.2044853 -0.1305869

> t.test(PSD[PGON==0], PSD[PGON==1])

     Welch Two Sample t-test

data:  PSD[PGON == 0] and PSD[PGON == 1]
t = -4.4065, df = 157.61, p-value = 1.933e-05
alternative hypothesis: true difference in me
95 percent confidence interval:
 -0.18879895 -0.07193228
sample estimates:
mean of x mean of y
 1.952116  2.082482

> t.test(CPSD[PGON==0], CPSD[PGON==1])

     Welch Two Sample t-test

data:  CPSD[PGON == 0] and CPSD[PGON == 1]
```

APPENDIX-3:

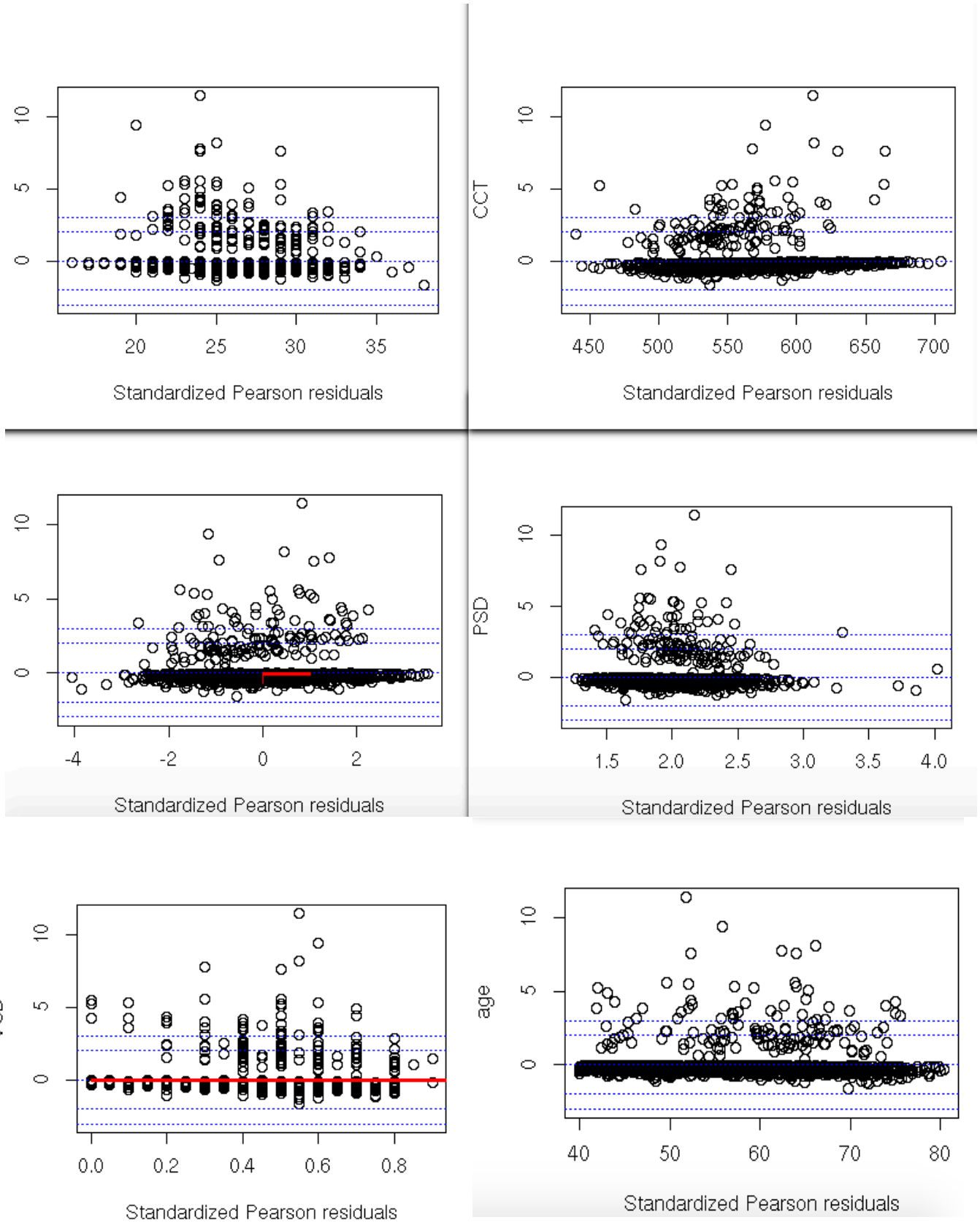
```
> summary(fit.uni)

Call:
glm(formula = PGON ~ Trt + race + gender + history + lbp + heart +
    hbp + diab + myopia + IOP + CCT + CD + VCD + PSD + MD + age,
    family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5875  -0.4116  -0.2555  -0.1558   3.0895

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.501884   2.091979  -2.152 0.031399 *
Trt         -1.175207   0.214037  -5.491 4.00e-08 ***
race        -0.182673   0.233810  -0.781 0.434633
gender       0.534236   0.202627   2.637 0.008376 **
history      0.321953   0.202637   1.589 0.112103
lbp          1.067585   0.406421   2.627 0.008619 **
heart        0.526876   0.344405   1.530 0.126063
hbp          0.345370   0.207457   1.665 0.095957 .
diab        -1.090403   0.452030  -2.412 0.015855 *
myopia       0.565271   0.216221   2.614 0.008940 **
IOP          0.161672   0.031880   5.071 3.95e-07 ***
CCT         -0.013071   0.002853  -4.581 4.63e-06 ***
CD          -0.453097   0.990404  -0.457 0.647321
VCD          3.199970   0.964572   3.318 0.000908 ***
PSD          1.025979   0.321535   3.191 0.001418 **
MD          -0.185908   0.089659  -2.074 0.038125 *
age          0.030644   0.011043   2.775 0.005519 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 951.79  on 1638  degrees of freedom
Residual deviance: 759.57  on 1622  degrees of freedom
AIC: 793.57

Number of Fisher Scoring iterations: 6

> AICc(fit.uni)
$AIC
[1] 793.5681

$AICc
[1] 793.9456

$BIC
[1] 885.3994
```

```
> summary(fit.best)

Call:
glm(formula = PGON ~ 1 + Trt + IOP + CCT + VCD + MD + PSD + age +
    gender + history + lbp + heart + diab + myopia, family = binom
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6279  -0.4152  -0.2574  -0.1560   3.1741

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.616511   2.007774  -2.299  0.02149 *
Trt         -1.176144   0.213816  -5.501 3.78e-08 ***
IOP          0.159872   0.031623   5.056 4.29e-07 ***
CCT         -0.012767   0.002747  -4.648 3.35e-06 ***
VCD          2.781316   0.522415   5.324 1.02e-07 ***
MD          -0.183753   0.088890  -2.067  0.03872 *
PSD          1.011458   0.316760   3.193  0.00141 **
age          0.032790   0.010857   3.020  0.00253 **
gender       0.521766   0.200196   2.606  0.00915 **
history      0.299911   0.201923   1.485  0.13747
lbp          0.984827   0.401465   2.453  0.01416 *
heart        0.581948   0.340736   1.708  0.08765 .
diab        -1.089184   0.451959  -2.410  0.01596 *
myopia       0.559901   0.214683   2.608  0.00911 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 951.79  on 1638  degrees of freedom
Residual deviance: 762.68  on 1625  degrees of freedom
AIC: 790.68

Number of Fisher Scoring iterations: 6

> AICc(fit.best)
$AIC
[1] 790.6808

$AICc
[1] 790.9394

$BIC
[1] 866.3066
> summary(backward)

Call:
glm(formula = PGON ~ Trt + IOP + CCT + VCD + MD + PSD + age
    gender + history + lbp + heart + hbp + diab + myopia, f
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5845  -0.4117  -0.2563  -0.1534   3.1236

Coefficients:
```

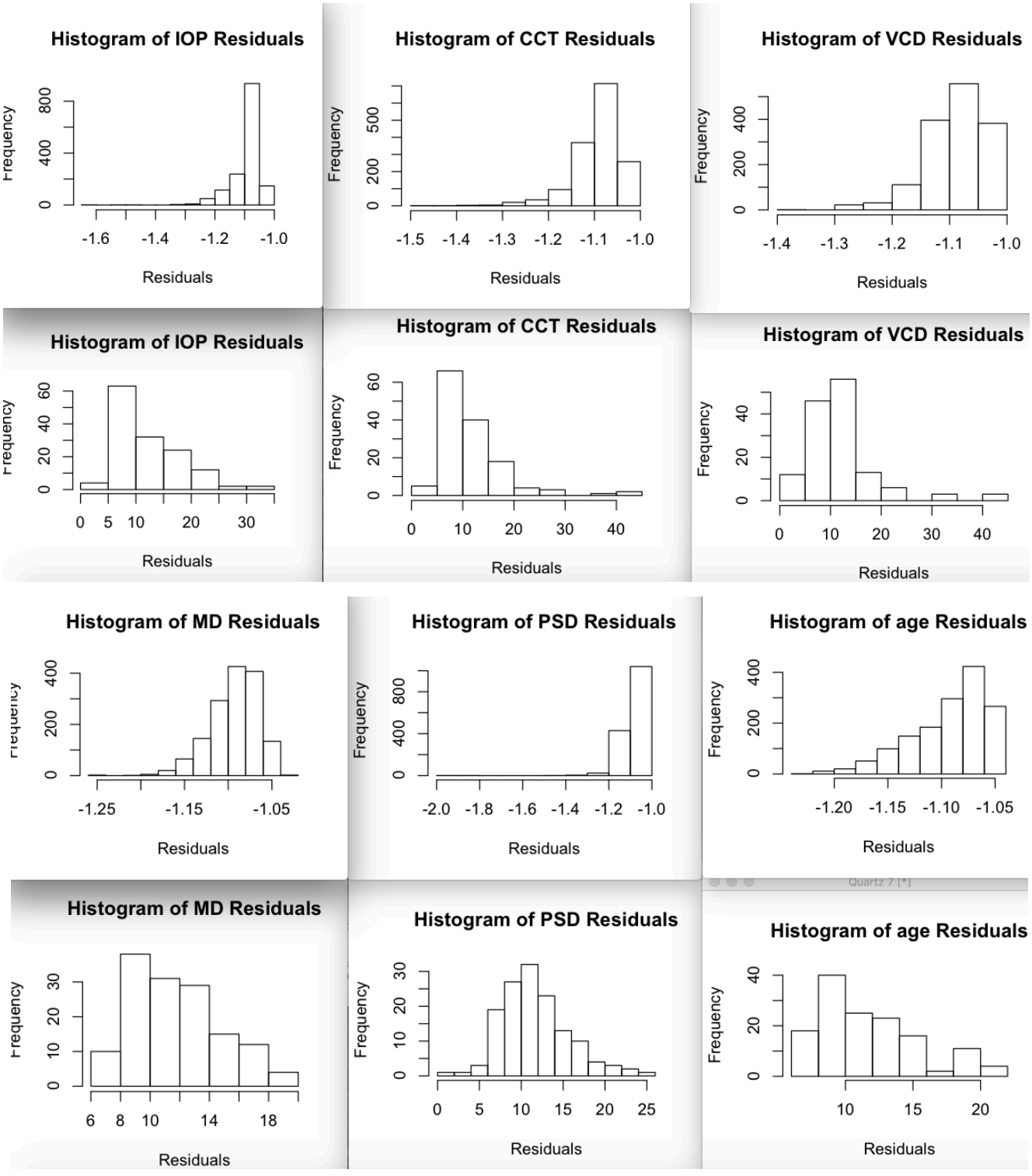
APPENDIX-4:


IOP
CCT
CD
VCD
MD
PSD
CPSD
age
450 650
0.0 0.6
1.5 3.0
40 60 80
20 35
450 650
0.0 0.6
0.0 0.6
-4 0
1.5 3.5
0.0 2.5
40 70
20 35
0.0 0.6
-4 0
0.0 2.0

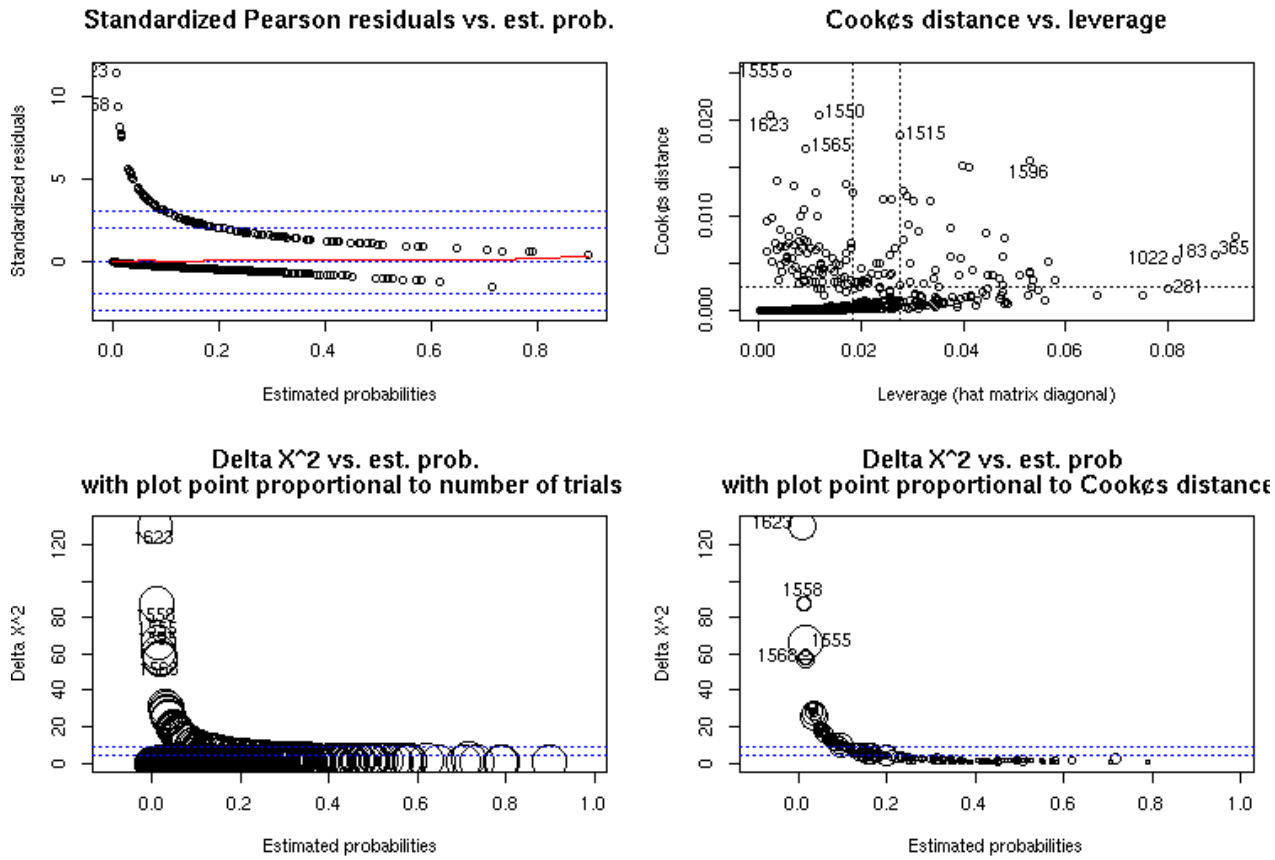
APPENDIX-5:


Standardized Pearson residuals
CCT
Standardized Pearson residuals
Standardized Pearson residuals
PSD
Standardized Pearson residuals
VCD
Standardized Pearson residuals
age
Standardized Pearson residuals

APPENDIX-6:


Histogram of IOP Residuals
Frequency
0
400
800
-1.6
-1.4
-1.2
-1.0
Residuals
Histogram of CCT Residuals
Frequency
0
200
500
-1.5
-1.4
-1.3
-1.2
-1.1
-1.0
Residuals
Histogram of VCD Residuals
Frequency
0
200
400
-1.4
-1.3
-1.2
-1.1
-1.0
Residuals
Histogram of IOP Residuals
Frequency
0
20
40
60
0
5
10
20
30
Residuals
Histogram of CCT Residuals
Frequency
0
20
40
60
0
10
20
30
40
Residuals
Histogram of VCD Residuals
Frequency
0
20
40
0
10
20
30
40
Residuals
Histogram of MD Residuals
Frequency
0
200
400
-1.25
-1.15
-1.05
Residuals
Histogram of PSD Residuals
Frequency
0
400
800
-2.0
-1.8
-1.6
-1.4
-1.2
-1.0
Residuals
Histogram of age Residuals
Frequency
0
200
400
-1.20
-1.15
-1.10
-1.05
Residuals
Histogram of MD Residuals
Frequency
0
10
20
30
6
8
10
14
18
Residuals
Histogram of PSD Residuals
Frequency
0
10
20
30
0
5
10
15
20
25
Residuals
Histogram of age Residuals
Frequency
0
20
40
10
15
20
Residuals

APPENDIX-7:



Deviance/df = 0.47; GOF thresholds: 2 SD = 1.07, 3 SD = 1.11

APPENDIX-8:

```
> HLTest(backward, 15)
$method
[1] "Hosmer and Lemeshow goodness-of-fit test with 15 bins"

$data.name
[1] "backward"

$statistic
        X2 parameter.df      p.value
  19.5402032   13.0000000    0.1072832

> lrtest(backward,Full)
Likelihood ratio test

Model 1: PGON ~ 1 + Trt + IOP + CCT + VCD + MD + PSD + age + gender +
    history + lbp + heart + hbp + diab + myopia
Model 2: PGON ~ eye + Trt + IOP + CCT + CD + VCD + MD + PSD + CPSD + ag
    race + gender + history + sysbb + ccb + migr + lbp + stroke +
    heart + hbp + diab + myopia
  #Df  LogLik Df Chisq Pr(>Chisq)
1  15 -380.19
2  23 -378.65  8 3.083      0.929
```

The code I used to obtain these results in R is as follows:

```
## LOAD DATASET ##
data<-read.table(file.choose("/Users/douglaswerre/Documents/DAEXAM/eye.csv"),header=T,sep="\t")

## LOAD PACKAGES ##
library(leaps)
library(MASS)
library(lmtest)
library(glmulti)

## DEFINE VARIABLES ##
ID<-data$Patient.ID
eye<-data$eye
PGON<-data$PGON
Trt<-data$Trt
IOP<-data$IOP
CCT<-data$CCT
CD<-data$CD
VCD<-data$VCD
MD<-data$MD
PSD<-data$PSD
CPSD<-data$CPSD
age<-data$age
race<-data$race
gender<-data$gender
history<-data$history
sysbb<-data$sysbb
ccb<-data$ccb
migr<-data$migr
lbp<-data$lbp
stroke<-data$stroke
heart<-data$heart
hbp<-data$hbp
diab<-data$diab
myopia<-data$myopia

 ## FIX VARIABLE 'EYE' ##
eye<-as.numeric(eye)
eye[eye==1] <- 0
eye[eye==2] <- 1
data$eye <- eye

data<-cbind(data[, 2:24])

# Univariate Analysis #
oddsratioWald.proc <- function(n00, n01, n10, n11, alpha = 0.05){

# Compute the odds ratio between two binary variables, x and y,
# as defined by the four numbers nij:
```

```
#      n00 = number of cases where x = 0 and y = 0
#      n01 = number of cases where x = 0 and y = 1
#      n10 = number of cases where x = 1 and y = 0
#      n11 = number of cases where x = 1 and y = 1
OR <- (n00 * n11)/(n01 * n10)

# Compute the Wald confidence intervals: #
siglog <- sqrt((1/n00) + (1/n01) + (1/n10) + (1/n11))
zalpha <- qnorm(1 - alpha/2)
logOR <- log(OR)
loglo <- logOR - zalpha * siglog
loghi <- logOR + zalpha * siglog
ORlo <- exp(loglo)
ORhi <- exp(loghi)
oframe <- data.frame(LowerCI = ORlo, OR = OR, UpperCI = ORhi, alpha = alpha)
oframe}
TABLEOR.proc <- function(x,y,alpha=0.05){xtab <- table(x,y)
n00 <- xtab[1,1]
n01 <- xtab[1,2]
n10 <- xtab[2,1]
n11 <- xtab[2,2]
outList <- vector("list",2)
outList[[1]] <- paste("Odds ration between the level [",dimnames(xtab)[[1]][1],"] of the first variable and the
 level [",dimnames(xtab) [[2]][2],"] of the second variable:",sep=" ")
outList[[2]] <- oddsratioWald.proc(n00,n01,n10,n11,alpha)
outList}

# Contingency Tables for Categorical Covariates with Odds Ratios #
table(PGON)
table(eye,PGON)
TABLEOR.proc(eye,PGON)
table(Trt,PGON)
TABLEOR.proc(Trt,PGON)
table(race,PGON)
TABLEOR.proc(race,PGON)
table(gender,PGON)
TABLEOR.proc(gender,PGON)
table(history,PGON)
TABLEOR.proc(history,PGON)
table(sysbb,PGON)
TABLEOR.proc(sysbb,PGON)
table(ccb,PGON)
TABLEOR.proc(ccb,PGON)
table(migr,PGON)
TABLEOR.proc(migr,PGON)
table(lbp,PGON)
TABLEOR.proc(lbp,PGON)
table(stroke,PGON)
TABLEOR.proc(stroke,PGON)
table(heart,PGON)
```

```
TABLEOR.proc(heart,PGON)
table(hbp,PGON)
TABLEOR.proc(hbp,PGON)
table(diab,PGON)
TABLEOR.proc(diab,PGON)
table(myopia,PGON)
TABLEOR.proc(myopia,PGON)

# 2-sample t-tests for continuous covariates #

t.test(IOP[PGON==0], IOP[PGON==1])
t.test(CCT[PGON==0], CCT[PGON==1])
t.test(CD[PGON==0], CD[PGON==1])
t.test(VCD[PGON==0], VCD[PGON==1])
t.test(MD[PGON==0], MD[PGON==1])
t.test(PSD[PGON==0], PSD[PGON==1])
t.test(CPSD[PGON==0], CPSD[PGON==1])
t.test(age[PGON==0], age[PGON==1])

# univariate logistic regression #

fit.eye=glm(PGON~eye, family=binomial(link=logit), data = data)
summary(fit.eye)
fit.Trt=glm(PGON~Trt, family=binomial(link=logit), data = data)
summary(fit.Trt)
fit.race=glm(PGON~race, family=binomial(link=logit), data = data)
summary(fit.race)
fit.gender=glm(PGON~gender, family=binomial(link=logit), data = data)
summary(fit.gender)
fit.history=glm(PGON~history, family=binomial(link=logit), data = data)
summary(fit.history)
fit.sysbb=glm(PGON~sysbb, family=binomial(link=logit), data = data)
summary(fit.sysbb)
fit.ccb=glm(PGON~ccb, family=binomial(link=logit), data = data)
summary(fit.ccb)
fit.migr=glm(PGON~migr, family=binomial(link=logit), data = data)
summary(fit.migr)
fit.lbp=glm(PGON~lbp, family=binomial(link=logit), data = data)
summary(fit.lbp)
fit.stroke=glm(PGON~stroke, family=binomial(link=logit), data = data)
summary(fit.stroke)
fit.heart=glm(PGON~heart, family=binomial(link=logit), data = data)
summary(fit.heart)
fit.hbp=glm(PGON~hbp, family=binomial(link=logit), data = data)
summary(fit.hbp)
fit.diab=glm(PGON~diab, family=binomial(link=logit), data = data)
summary(fit.diab)
fit.myopia=glm(PGON~myopia, family=binomial(link=logit), data = data)
summary(fit.myopia)
fit.IOP=glm(PGON~IOP, family=binomial(link=logit), data = data)
```

```
summary(fit.IOP)
fit.CCT=glm(PGON~CCT, family=binomial(link=logit), data = data)
summary(fit.CCT)
fit.CD=glm(PGON~CD, family=binomial(link=logit), data = data)
summary(fit.CD)
fit.VCD=glm(PGON~VCD, family=binomial(link=logit), data = data)
summary(fit.VCD)
fit.MD=glm(PGON~MD, family=binomial(link=logit), data = data)
summary(fit.MD)
fit.PSD=glm(PGON~PSD, family=binomial(link=logit), data = data)
summary(fit.PSD)
fit.CPSD=glm(PGON~CPSD, family=binomial(link=logit), data = data)
summary(fit.CPSD)
fit.age=glm(PGON~age, family=binomial(link=logit), data = data)
summary(fit.age)

######################### Multivariate Analysis ##########################

# I create a function to find the AIC, BIC, and AICc for each model #

AICc=function(object){
n=length(object$y)
r=length(object$coef)
AICc=AIC(object)+2*r*(r+1)/(n-r-1)
list(AIC=AIC(object), AICc=AICc, BIC=BIC(object))}


# Next I create a model using the variables that were significant in our univariate analysis #

fit.uni=glm(PGON~Trt+race+gender+history+lbp+heart+hbp+diab+myopia+IOP+CCT+CD+VCD+PSD+MD+
 age, family=binomial(link="logit"))
summary(fit.uni)
AICc(fit.uni)

# Now I will search through all possible models using all variables except for eye #

fit.all=glm(PGON~., family=binomial(link="logit"),data=data)
summary(fit.all)
search.aicc=glmulti(PGON~., data=data, fitfunction="glm", level=1, method="h", crit="aicc",
 family=binomial(link="logit"))
print(search.aicc)

aa=weightable(search.aicc)   # contains top 100 models with aicc #
cbind(model=aa[1:5, 1], round(aa[1:5,2:3], 3))
# weights are used for model averaging #

# best model found by AICc #

fit.best=glm(PGON~1+Trt+IOP+CCT+VCD+MD+PSD+age+gender+history+lbp+heart+diab+myopia,
 family=binomial(link=logit),data=data)
```

```
summary(fit.best)
AICc(fit.best)

## Backward Elimination ##

Base <- lm(PGON ~ -1, family = binomial(link = logit), data = data)
Full <- glm(PGON ~., family = binomial(link = logit), data = data)
summary(Full)
backward <- step(Full, direction = "backward", trace = FALSE)
summary(backward)
AICc(backward)
anova(backward)
plot(backward)
hist(backward$residuals, xlab = "Residuals", main = "Histogram of Residuals")

############################ WHICH IS BEST? ################################

Backward elimination gives us the best result

backward=glm(PGON~1+Trt+IOP+CCT+VCD+MD+PSD+age+gender+history+lbp+heart+hbp+diab+myopia,
 family=binomial(link=logit),data=data)

############################ INTERACTION EFFECTS ###########################

## We use a scatterplot matrix to look for interactions ##
plot(data[4:11])  # only looks at continuous variables #

## We expected to only use one of (CD,VCD) and PSD,(CPSD). This scatterplot shows the linear relationship
 between those variabels. ##

## First I will consider which variables might have interaction effects, and I include them in the model to test for
 significance, age*race seems most significant ##

plot(data[11:12])

fit.interactions=glm(PGON~1+Trt+IOP+CCT+VCD+MD+PSD+age+gender+history+lbp+heart+hbp+diab+my
 opia+age*race, family=binomial(link=logit),data=data)
summary(fit.interactions)
AICc(fit.interactions)

############################ TRANSFORMATIONS ###############################

## Histograms to see if we transform any explanatory variables ##

hist(fit.IOP$residuals[PGON==0], xlab = "Residuals", main = "Histogram of IOP Residuals")
quartz()
hist(fit.IOP$residuals[PGON==1], xlab = "Residuals", main = "Histogram of IOP Residuals")
quartz()
hist(fit.CCT$residuals[PGON==0], xlab = "Residuals", main = "Histogram of CCT Residuals")
quartz()
```

```
hist(fit.CCT$residuals[PGON==1], xlab = "Residuals", main = "Histogram of CCT Residuals")
quartz()
hist(fit.VCD$residuals[PGON==0], xlab = "Residuals", main = "Histogram of VCD Residuals")
quartz()
hist(fit.VCD$residuals[PGON==1], xlab = "Residuals", main = "Histogram of VCD Residuals")
quartz()
hist(fit.MD$residuals[PGON==0], xlab = "Residuals", main = "Histogram of MD Residuals")
quartz()
hist(fit.MD$residuals[PGON==1], xlab = "Residuals", main = "Histogram of MD Residuals")
quartz()
hist(fit.PSD$residuals[PGON==0], xlab = "Residuals", main = "Histogram of PSD Residuals")
quartz()
hist(fit.PSD$residuals[PGON==1], xlab = "Residuals", main = "Histogram of PSD Residuals")
quartz()
hist(fit.age$residuals[PGON==0], xlab = "Residuals", main = "Histogram of age Residuals")
quartz()
hist(fit.age$residuals[PGON==1], xlab = "Residuals", main = "Histogram of age Residuals")

#############################################################################
## FINAL MODEL AFTER TRANSFORMATIONS AND INTERACTIONS ##

## Consider transforming IOP, CCT, VCD, MD, PSD, and age ##

## Consider transforming IOP, CCT, VCD, MD, PSD, and age ##

## transform data into binomial (Convert to EVP form) ##
w <- aggregate(formula =
 PGON~1+Trt+IOP+CCT+VCD+MD+PSD+age+gender+history+lbp+heart+hbp+diab+myopia, data = data,
 FUN = sum)
n <- aggregate(formula =
 PGON~1+Trt+IOP+CCT+VCD+MD+PSD+age+gender+history+lbp+heart+hbp+diab+myopia, data = data,
 FUN = length)
w.n <- data.frame(w, trials = n$PGON, prop = round(w$PGON/n$PGON,2))
head(w.n)
nrow(w.n)  # Number of EVPs #
sum(w.n$trials)  # Number of observations
# Verify model fit to EVP data matches the model fit to the binary response data format #
mod.prelim1 <- glm(formula = PGON/trials ~
 1+Trt+IOP+CCT+VCD+MD+PSD+age+gender+history+lbp+heart+hbp+diab+myopia, family =
 binomial(link=logit), data = w.n, weights = trials)
round(summary(mod.prelim1)$coefficients, digits = 4)
round(summary(backward)$coefficients, digits = 4)

# Standardized residuals vs. IOP #
x11(width = 7, height = 6, pointsize = 12)
# pdf(file = "c\\figures\\Figure5.11color.pdf", width = 7, height = 6, color model = "cmyk")              # Create
 plot for book #
stand.resid <- rstandard(model = mod.prelim1, type = "pearson")
plot(x = w.n$IOP, y = stand.resid, ylim = c(min(-3, stand.resid), max(3, stand.resid)), xlab = "Standardized
 Pearson residuals", ylab = "IOP")
```

```
abline(h = c(3,2,0,-2,-3), lty = "dotted", col = "blue")
ord.IOP <- order(w.n$IOP)
smooth.stand <- loess(formula = stand.resid ~ IOP, data = w.n, weights = trials)
lines(x = w.n$PGON[ord.IOP], y = predict(smooth.stand)[ord.IOP], lty = "solid", col = "red")

# Standardized residuals vs. CCT #
x11(width = 7, height = 6, pointsize = 12)
# pdf(file = "c\\figures\\Figure5.11color.pdf", width = 7, height = 6, color model = "cmyk")              # Create
 plot for book #
stand.resid <- rstandard(model = mod.prelim1, type = "pearson")
plot(x = w.n$CCT, y = stand.resid, ylim = c(min(-3, stand.resid), max(3, stand.resid)), xlab = "Standardized
 Pearson residuals", ylab = "CCT")
abline(h = c(3,2,0,-2,-3), lty = "dotted", col = "blue")
ord.CCT <- order(w.n$CCT)
smooth.stand <- loess(formula = stand.resid ~ CCT, data = w.n, weights = trials)
lines(x = w.n$PGON[ord.CCT], y = predict(smooth.stand)[ord.CCT], lty = "solid", col = "red")

# Standardized residuals vs. VCD #
x11(width = 7, height = 6, pointsize = 12)
# pdf(file = "c\\figures\\Figure5.11color.pdf", width = 7, height = 6, color model = "cmyk")              # Create
 plot for book #
stand.resid <- rstandard(model = mod.prelim1, type = "pearson")
plot(x = w.n$VCD, y = stand.resid, ylim = c(min(-3, stand.resid), max(3, stand.resid)), xlab = "Standardized
 Pearson residuals", ylab = "VCD")
abline(h = c(3,2,0,-2,-3), lty = "dotted", col = "blue")
ord.VCD <- order(w.n$VCD)
smooth.stand <- loess(formula = stand.resid ~ VCD, data = w.n, weights = trials)
lines(x = w.n$PGON[ord.VCD], y = predict(smooth.stand)[ord.VCD], lty = "solid", col = "red")

# Standardized residuals vs. MD #
x11(width = 7, height = 6, pointsize = 12)
# pdf(file = "c\\figures\\Figure5.11color.pdf", width = 7, height = 6, color model = "cmyk")              # Create
 plot for book #
stand.resid <- rstandard(model = mod.prelim1, type = "pearson")
plot(x = w.n$MD, y = stand.resid, ylim = c(min(-3, stand.resid), max(3, stand.resid)), xlab = "Standardized
 Pearson residuals", ylab = "MD")
abline(h = c(3,2,0,-2,-3), lty = "dotted", col = "blue")
ord.MD <- order(w.n$MD)
smooth.stand <- loess(formula = stand.resid ~ MD, data = w.n, weights = trials)
lines(x = w.n$PGON[ord.MD], y = predict(smooth.stand)[ord.MD], lty = "solid", col = "red")

# Standardized residuals vs. PSD #
x11(width = 7, height = 6, pointsize = 12)
# pdf(file = "c\\figures\\Figure5.11color.pdf", width = 7, height = 6, color model = "cmyk")              # Create
 plot for book #
stand.resid <- rstandard(model = mod.prelim1, type = "pearson")
plot(x = w.n$PSD, y = stand.resid, ylim = c(min(-3, stand.resid), max(3, stand.resid)), xlab = "Standardized
 Pearson residuals", ylab = "PSD")
abline(h = c(3,2,0,-2,-3), lty = "dotted", col = "blue")
ord.PSD <- order(w.n$PSD)
```

```
smooth.stand <- loess(formula = stand.resid ~ PSD, data = w.n, weights = trials)
lines(x = w.n$PGON[ord.PSD], y = predict(smooth.stand)[ord.PSD], lty = "solid", col = "red")

# Standardized residuals vs. age #
x11(width = 7, height = 6, pointsize = 12)
# pdf(file = "c:\\figures\\Figure5.11color.pdf", width = 7, height = 6, color model = "cmyk")              # Create
 plot for book #
stand.resid <- rstandard(model = mod.prelim1, type = "pearson")
plot(x = w.n$age, y = stand.resid, ylim = c(min(-3, stand.resid), max(3, stand.resid)), xlab = "Standardized
 Pearson residuals", ylab = "age")
abline(h = c(3,2,0,-2,-3), lty = "dotted", col = "blue")
ord.age <- order(w.n$age)
smooth.stand <- loess(formula = stand.resid ~ age, data = w.n, weights = trials)
lines(x = w.n$PGON[ord.age], y = predict(smooth.stand)[ord.age], lty = "solid", col = "red")


##############################################################################################
                                        ## Diagnostic Plots ##
##############################################################################################
one.fourth.root=function(x){x^0.25}
#########################################################################
mod.fit.obj <- backward
scale.n = I
scale.cookd = I
pearson <- residuals(mod.fit.obj, type = "pearson")



 # Pearson residuals #

stand.resid <- rstandard(model = mod.fit.obj, type = "pearson")
# Standardized Pearson residuals #
deltaXsq <- stand.resid^2
pred <- mod.fit.obj$fitted.values
n <- mod.fit.obj$prior.weights       # Number of observations per EVP #
df <- mod.fit.obj$df.residual
cookd <- cooks.distance(mod.fit.obj)
h <- hatvalues(mod.fit.obj)
dev.res <- residuals(mod.fit.obj, type = "deviance")
stand.dev.resid <- rstandard(model = mod.fit.obj, type = "deviance")    # Standardized deviance residuals #
deltaD <- dev.res^2 + h*stand.resid^2
pear.stat <- sum(pearson^2)
dev <- mod.fit.obj$deviance
p <- length(mod.fit.obj$coefficients)

# Type of residuals to include on plots #
resid.plot11 <- stand.resid
resid.plot21 <- deltaXsq
plot.label11 <- "Pearson"
plot.label21 <- "Delta X^2"
```

```
##############################################################################

 # Four plots #
# Open a new plotting window
x11(width = 8, height = 6, pointsize = 12)
# Divide the plot into three rows and two columns. The last row is only 1cm in height to make sure there is some
 room for the printed GOF statistics
layout(mat = matrix(c(1,2,3,4,5,5), byrow = TRUE, ncol = 2), height = c(1,1,lcm(1)))
# layout.show(5)
# Standardized residual vs predicted prob.
plot(x = pred, y = resid.plot11, xlab = "Estimated probabilities", ylab = "Standardized residuals", main =
 paste("Standardized", plot.label11, "residuals vs. est. prob."), ylim = c(min(-3, stand.resid), max(3,
 stand.resid)))
abline(h = c(-3, -2, 0, 2, 3), lty = "dotted", col = "blue")
order.pred <- order(pred)
smooth.stand <- loess(formula = resid.plot11 ~ pred, weights = n)
lines(x = pred[order.pred], y = predict(smooth.stand)[order.pred], lty = "solid", col = "red")

identify(x = pred, y = resid.plot11, labels = labels(pred))

# Cook's distance vs leverage #
plot(x = h, y = cookd, ylim = c(0, max(4/length(cookd), cookd)), xlim = c(0, max(3*p/length(h), h)), xlab =
 "Leverage (hat matrix diagonal)", ylab = "Cook's distance", main = "Cook's distance vs. leverage")
abline(h = c(4/length(cookd), 1), lty = "dotted")
abline(v = c(2*p/length(h), 3*p/length(h)), lty = "dotted")

identify(x = h, y = cookd, labels = labels(h))

# Delta X^2 or D vs. prediction prob. with plotting point proportional to n
symbols(x = pred, y = resid.plot21, xlab = "Estimated probabilities", circles = scale.n(n), ylab = plot.label21,
 main = paste(plot.label21, "vs. est. prob. \n with plot point proportional to number of trials"), inches = 0.1, ylim
 = c(0, max(9, resid.plot21)))
abline(h = c(4, 9), lty = "dotted", col = "blue")

identify(x = pred, y = resid.plot21, labels = labels(pred))


# Print deviance/df on plot
dev.df <- dev/df
gof.threshold <- round(c(1 + 2*sqrt(2/df), 1 + 3*sqrt(2/df)), 2)
mtext(text = paste("Deviance/df =  ", round(dev.df, 2),  "; GOF thresholds: 2 SD =  ",round(gof.threshold[1], 2),
 ", 3 SD = ",round(gof.threshold[2], 2), sep = ""), side = 1, line = 6, cex = 1.0, adj = 0)

# Delta X^2 or D vs. predicted prob. with plotting point proportional to Cook's distance

symbols(x = pred, y = resid.plot21, circles = scale.cookd(cookd), xlab = "Estimated probabilities", ylab =
 plot.label21, main = paste(plot.label21, "vs. est. prob \n with plot point proportional to Cook's distance"), inches
 = 0.1, ylim = c(0, max(9, resid.plot21)))
```

```
abline(h = c(4, 9), lty = "dotted", col = "blue")

identify(x=pred, y = resid.plot21, labels = labels(pred))

save1 = examine.logistic.reg(mod.prelim1, identify.points = T, scale.n = one.fourth.root, scale.cookd = sqrt)}

########################################
# examine the EVPs identified in the plots #
########################################

pi.hat = round(pred, 2)
std.res = round(stand.resid, 2)
cookd = round(cookd, 2)
h = round(h, 2)

w.n.diag1 = cbind(w.n, pi.hat, std.res, cookd, h)

p = length(backward$coef)
ck.out=abs(w.n.diag1$std.res)>2 | w.n.diag1$cookd>4/nrow(w.n) | w.n.diag1$h > 3*p/nrow(w.n)
extract.EVPs = w.n.diag1[ck.out, ]
extract.EVPs[order(extract.EVPs$PGON),]     # order by PGON #

############################
# Hosmer-Lemeshow test #
############################
HLTest = function(obj, g)  {
stopifnot(family(obj)$family == "binomial" && family(obj)$link == "logit")
y = obj$model[[1]]
trials = rep(1, times = nrow(obj$model))
if(any(colnames(obj$model) == "(weights)"))
trials <- obj$model[[ncol(obj$model)]]
if (is.factor(y))
y = as.numeric == 2     # Converts 1-2 factor levels to logical 0/1 values #
yhat = obj$fitted.values
interval = cut(yhat, quantile(yhat, 0:g/g), include.lowest = TRUE)        # Creates factor with levels 1,2,..,g #
Y1 <- trials*y
Y0 <- trials - Y1
Y1hat <- trials*yhat
Y0hat <- trials - Y1hat
obs = xtabs(formula = cbind(Y0,Y1) ~ interval)
expect = xtabs(formula = cbind(Y0hat, Y1hat) ~ interval)
if (any(expect < 5))
pear <- (obs - expect)/sqrt(expect)
chisq = sum(pear^2)
P = 1 - pchisq(chisq, g - 2)
return(structure(list(method = c(paste("Hosmer and Lemeshow goodness-of-fit test with", g, "bins", sep = " ")),
 data.name = deparse(substitute(obj)), statistic = c(X2 = chisq, parameter = c(df = g-2), p.value = P)))){
HLTest(backward, 15)
# good overall fit #
```

##

lrtest(backward,Full)

DONE