Report for Choosing Appropriate Logistic Regression Model

A. **Table 1:** Adjusted association between hippocampal sclerosis and Global brain arteriolosclerosis in three logistic regression models (n =1215)

Model	Variable	Adjusted Odds Ratio	95% Wald CI for Odds ratio
Proportional	HS (Yes vs No)	0.88	(0.60, 1.30)
Odds	Gender (Female vs Male)	0.72	(0.72, 1.12)
Model	Age (1 year)	0.93	(0.91, 0.94)
Partial	HS (Yes vs No) for No GBA*	0.24	(0.09,0.70)
Proportional	HS (Yes vs No) for Mild GBA**	1.1	(0.70,1.65)
Odds	HS (Yes vs No) for Moderate GBA***	1.82	(0.65,5.13)
Model	Gender (Female vs Male) for No GBA*	0.7	(0.49,0.91)
	Gender (Female vs Male) for Mild GBA**	1	(0.78,1.28)
	Gender (Female vs Male) for Moderate GBA***	1.6	(1.00,2.61)
	Age (1 year)	0.93	(0.91,0.94)
Multinomial	HS (Yes vs No) for No GBA***	0.6	(0.13,2.31)
Model	HS (Yes vs No) for Mild GBA***	2.41	(0.83,7.01)
	HS (Yes vs No) for Moderate GBA***	2.1	(0.71, 6.20)
	Gender (Female vs Male) for No GBA***	1.22	(0.70, 2.20)
	Gender (Female vs Male) for Mild GBA***	1.85	(1.10,3.11)
	Gender (Female vs Male) for Moderate GBA***	1.9	(1.12,3.23)
	Age (1 year) for No GBA***	0.83	(0.80,0.90)
	Age (1 year) for Mild GBA***	0.9	(0.84,0.92)
	Age (1 year) for Moderate GBA***	0.91	(0.90,1.00)
Note:	HS = hippocampal sclerosis		
	GBA = Global brain arteriolosclerosis		
	* compared to Any Global brain arteriolosclerosis		
	** compared to More severe Global brain arteriolosclerosis		
	*** compared to Severe Global brain arteriolosclerosis		

B. **Results:**

For proportional odds model, compared to those not having hippocampal sclerosis (HS), those having HS have 12% reduce of reporting less severe Global brain arteriolosclerosis (GBA), for no GBA compared to mild, moderate or severe GBA; for no or mild GBA compared to moderate or severe GBA; and for no, mild or moderate GBA compared to severe GBA, assuming other variables are constant.

For partial proportional odds model, compared to those not having HS, those having HS have 76% reduced odds of no GBA compared to any GBA; 10% increased odds of no or mild GBA compared to more severe GBA; 82% increased odds of moderate or less severe up to moderate compared to severe GBA, assuming other variables are constant.

For multinomial model, compared to those not having HS, those having HS have 0.6 times the odds of no GBA compared to severe GBA; 2.41 times the odds of mild GBA compared to severe GBA; 2.1 the odds of moderate GBA compared to severe GBA, assuming other variables are constant.

Choosing model:

- Based on AIC, multinomial model seems to be the best since it has lowest AIC.
- We would look at the simplest model, using fewest number of variables first, because more variables need more data, the estimates may be wrong. In this case, we start with the proportional odds. Since the score test for proportional odds assumption here has $p<0.0001$, the proportional odds assumption is not hold. So we move to other 2 models.
- Looking at the non- proportional odds cumulative model, linear hypothesis testing result table, both HS and female have p values <0.05 , but not age at death, which means we should estimate a parameter for HS for each logit, for gender for each logit, and proportional odds assumption for age. This shows that partial proportion is preferred compared to proportional odds.
- For the 3rd model multinomial model, it is preferred when the outcome is nominal however, it's obvious that our outcome is ordinal, so I think multinomial would not be a good choice to answer our research question.
- Therefore, the best model out of all 3 would be partial proportional odds.

Strength and limitation, discrepancies and similarities:

- The strengths for proportional odds and partial proportional odds are (1) both consider proportionality assumption (2) both treat outcome as ordinal variable, which is related to our research question, (3) partial proportional odds model, not proportional odds model, accounts for failing proportionality assumption for regression parameters, which is part of our situation here, (4) both use fewer number of variables- proportional odds uses fewer than partial proportional odds. The limitation is (1) both of them have higher AIC compared to multinomial model, and (2) they might not be good choices if there is truly no proportionality present.
- The strength for multinomial model is (1) having higher goodness of fit compared to other 2 models while the limitation is treating outcome as nominal variable which is not similar to our research question.
- Looking at parameter estimate for each logit for proportional odds model, they are -0.13, -0.11, -0.08 for HS, female and age at death respectively, which is pretty close to each other. But for partial proportion model, we could see differences in sign (positive vs negative) for HS and female logits. And for multinomial model, for HS, again we see difference in sign, difference in parameter estimation values for female, and pretty similar result for age at death.

C. **Appendix:**

****PROPORTIONAL ODDS MODEL;**

```
proc logistic data=_temp0.arter;
class hs(ref="0") female(ref="0")/param=ref;
model arter=hs female npdage/aggregate scale=none;
run;
```

The LOGISTIC Procedure

Model Information	
Data Set	_TEMP0.ARTER
Response Variable	arter
Number of Response Levels	4
Model	cumulative logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1225
Number of Observations Used	1215

Response Profile

Ordered Value	arter	Total Frequency
1	1	205
2	2	560
3	3	380
4	4	70

Probabilities modeled are cumulated over the lower Ordered Values.

Note: 10 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information

Class	Value	Design Variables
HS	**0**	0
	1	1
Female	**0**	0
	1	1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
32.8306	6	<.0001

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	2598.8030	3312	0.7847	1.0000
Pearson	3194.8309	3312	0.9646	0.9263

Number of unique profiles: 1106

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	2886.031	2799.165
SC	2901.338	2829.780
-2 Log L	2880.031	2787.165

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	92.8659	3	<.0001
Score	88.3119	3	<.0001
Wald	90.4380	3	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
HS	1	0.4314	0.5113
Female	1	0.8791	0.3484
NPDAGE	1	82.2521	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	5.0570	0.7162	49.8589	<.0001
Intercept	2	1	7.3033	0.7339	99.0233	<.0001
Intercept	3	1	9.6412	0.7535	163.6986	<.0001
HS	1	1	-0.1300	0.1980	0.4314	0.5113
Female	1	1	-0.1057	0.1127	0.8791	0.3484
NPDAGE		1	-0.0750	0.00827	82.2521	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
HS 1 vs 0	0.878	0.596	1.294

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Female 1 vs 0	0.900	0.721	1.122
NPDAGE	0.928	0.913	0.943

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	62.3	Somers' D	0.254
Percent Discordant	36.9	Gamma	0.255
Percent Tied	0.8	Tau-a	0.167
Pairs	485650	c	0.627

```
**PARTIAL-PROPORTIONAL ODDS MODEL;
proc logistic data=_temp0.arter outest=sas;
class hs(ref="0") female(ref="0")/param=ref;
model arter=hs female npdage/unequalslopes aggregate scale=none;
hs: test HS1_2=HS1_1;
hs: test HS1_3=HS1_2;
female: test female1_2=female1_1;
female: test female1_3=female1_2;
npdage: test npdage_2=npdage_1;
npdage: test npdage_3=npdage_1;
run;
```

The LOGISTIC Procedure

Model Information	
Data Set	_TEMP0.ARTER
Response Variable	arter
Number of Response Levels	4
Model	cumulative logit
	non-proportional odds
Optimization Technique	Newton-Raphson ridge

Number of Observations Read	1225
Number of Observations Used	1215

Response Profile		
Ordered Value	arter	Total Frequency
1	1	205
2	2	560
3	3	380
4	4	70

Probabilities modeled are cumulated over the lower Ordered Values.

Note: 10 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information

Class	Value	Design Variables
HS	0	0
	1	1
Female	0	0
	1	1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	2562.4476	3306	0.7751	1.0000
Pearson	3270.8593	3306	0.9894	0.6648

Number of unique profiles: 1106

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	2886.031	2774.809
SC	2901.338	2836.039
-2 Log L	2880.031	2750.809

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	129.2213	9	<.0001
Score	125.3910	9	<.0001
Wald	118.3126	9	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
HS	3	9.9414	0.0191
Female	3	13.4272	0.0038
NPDAGE	3	93.3755	<.0001

Analysis of Maximum Likelihood Estimates							
Parameter		arter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1	5.8266	0.9993	33.9950	<.0001
Intercept		2	1	6.3286	0.8378	57.0657	<.0001
Intercept		3	1	13.7878	1.8302	56.7538	<.0001
HS	1	1	1	-1.4261	0.5209	7.4951	0.0062
HS	1	2	1	0.0386	0.2191	0.0310	0.8602
HS	1	3	1	0.7636	0.5308	2.0690	0.1503
Female	1	1	1	-0.4027	0.1589	6.4267	0.0112
Female	1	2	1	-0.0317	0.1275	0.0617	0.8038
Female	1	3	1	0.5959	0.2530	5.5497	0.0185
NPDAGE		1	1	-0.0814	0.0116	49.0490	<.0001
NPDAGE		2	1	-0.0648	0.00945	47.0784	<.0001
NPDAGE		3	1	-0.1253	0.0198	39.8731	<.0001

Odds Ratio Estimates				
Effect	arter	Point Estimate	95% Wald Confidence Limits	
HS 1 vs 0	1	0.240	0.087	0.667
HS 1 vs 0	2	1.039	0.676	1.597
HS 1 vs 0	3	2.146	0.758	6.074
Female 1 vs 0	1	0.668	0.490	0.913
Female 1 vs 0	2	0.969	0.755	1.244
Female 1 vs 0	3	1.815	1.105	2.979
NPDAGE	1	0.922	0.901	0.943
NPDAGE	2	0.937	0.920	0.955
NPDAGE	3	0.882	0.849	0.917

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	62.4	Somers' D	0.255
Percent Discordant	36.9	Gamma	0.257
Percent Tied	0.7	Tau-a	0.168
Pairs	485650	c	0.627

Linear Hypotheses Testing Results			
Label	Wald Chi-Square	DF	Pr > ChiSq
hs	7.5693	1	0.0059
hs	1.9189	1	0.1660
female	4.9050	1	0.0268
female	6.6082	1	0.0102
npdage	1.7493	1	0.1860
npdage	3.9543	1	0.0468

**based on linear hypothesis testing, hs and female do not appear to have proportionality across all logits, but age does so final model will have unequal slopes for hs and female, and proportional odds assumption for age;

```
proc logistic data=_temp0.arter;
class hs(ref="0") female(ref="0")/param=ref;
model arter=hs female npdage/unequalslopes=(hs female) aggregate scale=none;
run;
```

The LOGISTIC Procedure

Model Information	
Data Set	_TEMP0.ARTER
Response Variable	arter
Number of Response Levels	4
Model	cumulative logit
	partial proportional odds
Optimization Technique	Newton-Raphson ridge

Number of Observations Read	1225
Number of Observations Used	1215

Response Profile		
Ordered Value	arter	Total Frequency
1	1	205
2	2	560
3	3	380
4	4	70

Probabilities modeled are cumulated over the lower Ordered Values.

Note: 10 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information		
Class	Value	Design Variables
HS	0	0

Class Level Information		
Class	**Value**	**Design Variables**
	1	1
Female	**0**	0
	1	1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	**Value**	**DF**	**Value/DF**	**Pr > ChiSq**
Deviance	2573.6103	3308	0.7780	1.0000
Pearson	3208.3904	3308	0.9699	0.8904

Number of unique profiles: 1106

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	2886.031	2781.972
SC	2901.338	2832.997
-2 Log L	2880.031	2761.972

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	118.0585	7	<.0001
Score	113.3616	7	<.0001
Wald	106.3586	7	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
HS	3	9.4009	0.0244
Female	3	11.4981	0.0093
NPDAGE	1	81.5998	<.0001

Analysis of Maximum Likelihood Estimates							
Parameter		**arter**	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		**1**	1	5.3073	0.7267	53.3317	<.0001

Analysis of Maximum Likelihood Estimates							
Parameter		arter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		2	1	7.2457	0.7393	96.0562	<.0001
Intercept		3	1	9.2531	0.7663	145.8204	<.0001
HS	1	1	1	-1.4363	0.5202	7.6238	0.0058
HS	1	2	1	0.0686	0.2199	0.0974	0.7550
HS	1	3	1	0.5996	0.5284	1.2878	0.2565
Female	1	1	1	-0.4089	0.1582	6.6837	0.0097
Female	1	2	1	-0.00501	0.1277	0.0015	0.9687
Female	1	3	1	0.4695	0.2505	3.5125	0.0609
NPDAGE			1	-0.0753	0.00834	81.5998	<.0001

Odds Ratio Estimates				
Effect	arter	Point Estimate	95% Wald Confidence Limits	
HS 1 vs 0	1	0.238	0.086	0.659
HS 1 vs 0	2	1.071	0.696	1.648
HS 1 vs 0	3	1.821	0.647	5.130
Female 1 vs 0	1	0.664	0.487	0.906
Female 1 vs 0	2	0.995	0.775	1.278
Female 1 vs 0	3	1.599	0.979	2.613
NPDAGE		0.927	0.912	0.943

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	62.3	Somers' D	0.255
Percent Discordant	36.9	Gamma	0.257
Percent Tied	0.8	Tau-a	0.168
Pairs	485650	c	0.627

****multinomial model;**

```
proc logistic data=_temp0.arter ;
class hs(ref="0") female(ref="0")/param=ref;
model arter=hs female npdage/aggregate scale=none link=glogit;
run;
```

The LOGISTIC Procedure

Model Information	
Data Set	_TEMP0.ARTER
Response Variable	arter
Number of Response Levels	4

Model Information	
Model	generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	1225
Number of Observations Used	1215

Response Profile

Ordered Value	arter	Total Frequency
1	1	205
2	2	560
3	3	380
4	4	70

Logits modeled use arter=4 as the reference category.

Note: 10 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information

Class	Value	Design Variables
HS	**0**	0
	1	1
Female	**0**	0
	1	1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	2564.7736	3306	0.7758	1.0000
Pearson	3268.2767	3306	0.9886	0.6763

Number of unique profiles: 1106

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	2886.031	2777.135
SC	2901.338	2838.365
-2 Log L	2880.031	2753.135

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	126.8953	9	<.0001
Score	118.7674	9	<.0001
Wald	107.9595	9	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
HS	3	9.9171	0.0193
Female	3	11.6711	0.0086
NPDAGE	3	82.8482	<.0001

Analysis of Maximum Likelihood Estimates							
Parameter		**arter**	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		**1**	1	18.1331	2.1655	70.1171	<.0001
Intercept		**2**	1	13.5646	2.0113	45.4861	<.0001
Intercept		**3**	1	9.8895	2.0333	23.6553	<.0001
HS	**1**	**1**	1	-0.5983	0.7344	0.6637	0.4152
HS	**1**	**2**	1	0.8810	0.5442	2.6215	0.1054
HS	**1**	**3**	1	0.7374	0.5511	1.7903	0.1809
Female	**1**	**1**	1	0.2020	0.2917	0.4795	0.4887
Female	**1**	**2**	1	0.6128	0.2656	5.3245	0.0210
Female	**1**	**3**	1	0.6405	0.2710	5.5881	0.0181
NPDAGE		**1**	1	-0.1915	0.0238	64.4994	<.0001
NPDAGE		**2**	1	-0.1310	0.0219	35.8337	<.0001
NPDAGE		**3**	1	-0.0942	0.0221	18.1860	<.0001

Odds Ratio Estimates				
Effect	**arter**	**Point Estimate**	**95% Wald Confidence Limits**	
HS 1 vs 0	**1**	0.550	0.130	2.319
HS 1 vs 0	**2**	2.413	0.831	7.012
HS 1 vs 0	**3**	2.090	0.710	6.157
Female 1 vs 0	**1**	1.224	0.691	2.168
Female 1 vs 0	**2**	1.846	1.097	3.106
Female 1 vs 0	**3**	1.897	1.116	3.227

Odds Ratio Estimates				
Effect	arter	Point Estimate	95% Wald Confidence Limits	
NPDAGE	1	0.826	0.788	0.865
NPDAGE	2	0.877	0.840	0.916
NPDAGE	3	0.910	0.872	0.950