AA-31: Logistic Regression Odyssey on a Small Sample Due to Rare Cancer

Brandy is a statistician and database programmer at the University of Michigan Medical School. She has worked on a wide variety of research projects. She holds a bachelors degree in engineering, and masters degrees in mathematics and statistics. The American Public Health Association awarded Brandy a student research award for her statistics masters thesis on path analysis. She has also received "Advanced Analytics" and "Data for Good" awards from the Midwest SAS Users Group conference.

Outside of her UM job, Brandy is a pianist, vocalist, and composer, an assistant tai chi instructor, and also sings in choirs, and enjoys martial arts aerobics and yoga.

MWSUG 2025
Cincinnati, OH

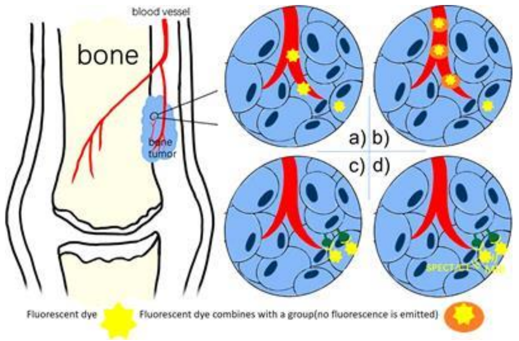
Logistic Regression Odyssey on a Small Sample Due to Rare Cancer

Brandy R. Sinco, MA, MS, University of Michigan, Ann Arbor, MI

Jessie M. Dalman, BA (Medical Student, MD Candidate)

Tasha M. Hughes, MD, MPH

Christina V. Angeles, MD



Background for Project

- Sarcoma is a cancer arising from bone and connective tissue (fat, muscle, blood vessels, nerves and the tissue that surrounds bones and joints). 150+ subtypes of sarcoma.
- Leiomyosarcoma (LMS) is a rare and aggressive form of sarcoma. LMS originates in the smooth muscle cells of organs like the uterus, stomach, intestines, and blood vessels.
- Due to the rare disease, the data sample was small. Original dataset had N = 116 patients. After data cleaning, N = 115 patients, considered large for a sarcoma study.
- Researchers hypothesized that patients with sub-cutaneous extensions will have higher odds of recurrence.
- Due to small sample and <80% power to detect a difference in recurrence at p<.05, originally thought that Bayesian analysis would be the best option.
- Bayesian analysis does not rely on p-values nor on asymptotics.

Outline

1)_Pearson Chi-Square and Fisher Exact Test.

2)_Power calculations with Proc Power.

3)_Classical Logistic Regression – Explanation of the Firth Adjustment.

4)_Classical Logistic Regression SAS code, output, comments.

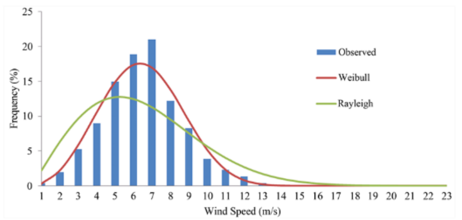
5)_Introduction to Bayesian Analysis.

6)_Bayesian Logistic Regression with Procs GenMod and BGLIMM.

7)_Conclusions and Closing Comments

Statistical Power (1 of 3)

- Power to detect a significant effect when it is present.
- Example: Detect the presence of mice in a house, based on chemical present on the bodies of mice.
- Cats and dogs have more power to detect the presence of some chemicals than humans (example: bomb sniffing dogs).
- Power depends on the minimum amount to detect, variation in the data, sample size.
- Detecting chemical concentration of 100 would require a smaller sample size than if we wanted to detect 10.
- Smaller differences require larger sample sizes to detect. Example: Compare proportions of college-educated people between two college towns would require larger sample size than between college town and rural community.



Statistical Power (2 of 3)

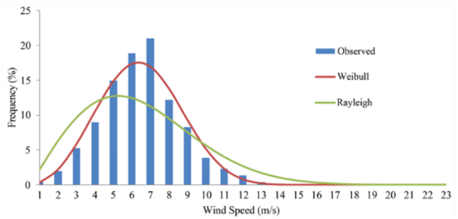
- μ = Parameter of interest, a mean, variance, rate, coefficient, or regression curve.

- Power = Pr(Detecting difference in μ | Difference in μ is present).

- Example: Compare tumor recurrence rates between patients with cutaneous-only versus cutaneous + sub-cutaneous extension.

- H_0: Null hypothesis, no difference, $\mu_1 = \mu_{2.}$
- No impact of tumor sub-type on recurrence.

- H_A: Alternative hypothesis, $\mu_1 \neq \mu_2$ or $\mu_1 > \mu_{2.}$ or $\mu_1 < \mu_{2.}$
- Recurrence rate higher based on tumor sub-type.



Statistical Power (3 of 3)

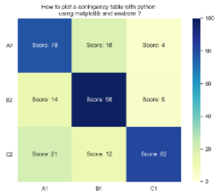
- Type 1 error, α = Pr(H_A | H_0 is true). α also called “level of significance”.

- Type 1 error: Conclude that a difference is present, when difference is due to random variation.
- I.E., wrongfully conclude that recurrence differs by tumor sub-type, when there is really no difference.

- P-value = observed level of significance.

- Type 2 error, β = Pr(H_0 | H_A). Fail to detect a difference when a difference is present.

➢ **Power = 1 – β = Pr(H_A | H_A). Probability of accurately detecting a difference.**



Pearson Chi-Square and Fisher Exact Tests

- Both tests used on contingency tables to test for significant differences between groups, such as tumor sub-types.

- Pearson chi-square test is computed based on comparison expected cell counts and observed cell counts.

- When expected cell counts are all ≥ 5, the Pearson X^2 is usually the better choice because the computation is faster.

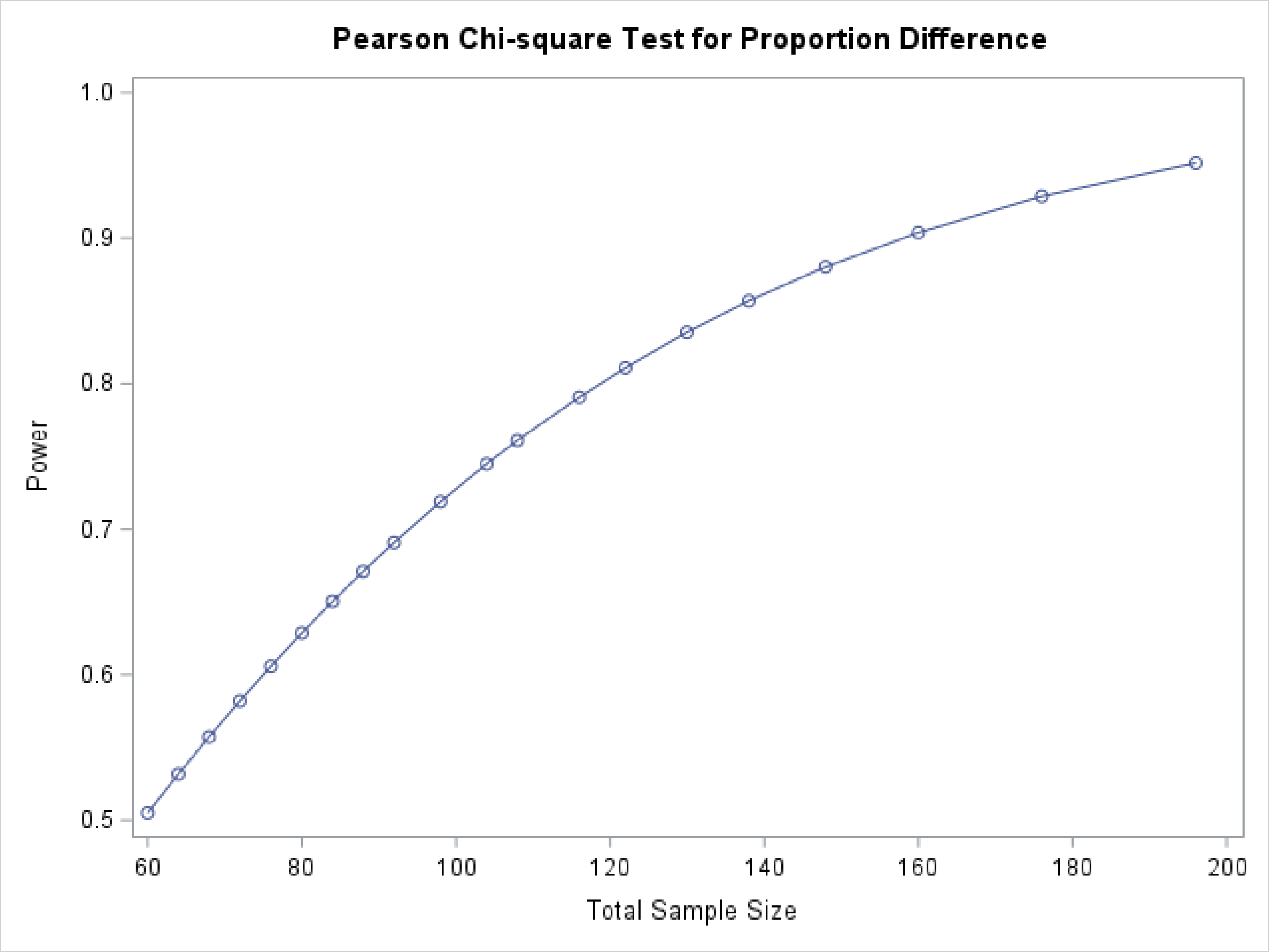
- When SAS gives the warning in the log, “cells have expected counts less than 5. Chi-Square may not be a valid test”, user Fisher’s exact test, which is based on a hyper-geometric distribution.

SAS Code for Power

```
/* Pearson Chi-Square Test */
Proc Power;
TwoSampleFreq test=pchi alpha=.05    /* Use test=fisher for Fisher exact test */
  groupproportions = (.02 .12)
  groupweights= (1 1)
  ntotal=.
  power= .5 to .95 by .05;
Plot Y=power;
Run;
```

Result: <60% power to detect a 10% difference at p<.05 between 2% and 12% recurrence. For the 80% power that is normally recommended, we would need 204 patients.

Power Plot from Proc Power



Recurrence by Sarcoma Sub-Type – Fisher Exact Test

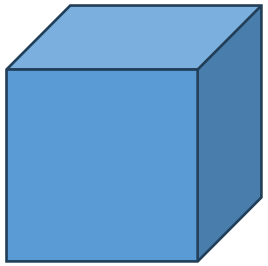
- Original dataset had N = 116 patients.

Recurrence Location	**Cutaneous (dermal) only** N = 72	**Cutaneous with subcutaneous extension** N = 44	**P-Value**
Total Recurrence	3 (4.2%)	6 (13.6%)	0.081
Local only	0	3 (7.7%)	0.052
Distant only	3 (3.8%)	3 (7.7%)	0.399

Recurrence by Sarcoma Sub-Type – Fisher Exact Test

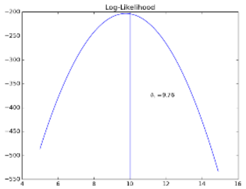
- Corrected dataset had N = 115 patients (1 excluded, 1 sub-cut ➔ cutaneous)

Recurrence Location	Cutaneous (dermal) only N = 70	Cutaneous with subcutaneous extension N = 45	P-Value
Total Recurrence	1 (1.4%)	7 (15.6%)	0.006
Local only	0 (0.0%)	3 (6.7%)	0.058
Distant only	1 (1.4%)	2 (4.4%)	0.560



Surprise!

- Even though power was low, statistically significant difference in recurrence by sub-type, based on the Fisher exact test.
- In a dataset with a categorical outcome, removing or re-categorizing a small number of observations can make a huge difference.
- More surprises to follow.

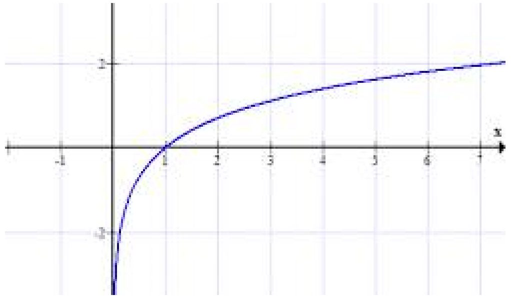


FIRTH MAXIMUM LIKELIHOOD, SEPARATION, 1 of 6

- Logistic regression coefficients are calculated from maximum likelihood estimation (mle), provided that the likelihood function has a unique maximum.

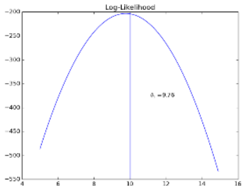
- Complete separation means that a single variable or a linear combination of variables perfectly predicts the outcome. IE, the outcome variable perfectly separates a predictor variable.
- Quasi-complete separation occurs when there is complete separation, except for a single observation of a predictor variable.
- Complete and quasi-separation are risks for small datasets.

- Example of “Complete Separation” from /stats.oarc.ucla.edu.
- Y = 0 all have values of X1<=3 and observations with Y = 1 all have values of X1>3.



FIRTH MAXIMUM LIKELIHOOD, SEPARATION, 2 of 6

- When separation occurs, the mle function does not have a unique maximum.
- data t; input Y X1 X2; datalines;   0 1 3
- 0 2 2
- 0 3 -1
- 0 3 -1
- 1 5 2
- 1 6 4
- 1 10 1
- 1 11 0; run;
- proc logistic data = t descending;
-   model y = x1 x2;   run;

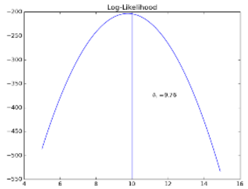


FIRTH MAXIMUM LIKELIHOOD, SEPARATION, 3 of 6

- Firth’s penalized maximum likelihood function can sometimes provide a maximum likelihood estimate, when the dataset has a separation problem.
- Firth, D. (1993). “Bias Reduction of Maximum Likelihood Estimates.” Biometrika 80:27–38.

- Original Output:
- Complete separation of data points detected. Warning: The maximum likelihood estimate does not exist.

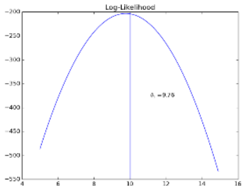
Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	89.311	<0.001	>999.999
X2	10.980	<0.001	>999.999



FIRTH MAXIMUM LIKELIHOOD, SEPARATION, 4 of 6

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-21.4410	72.9133	0.0865	0.7687
X1	1	4.8253	16.8543	0.0820	0.7747
X2	1	2.1690	23.9528	0.0082	0.9278

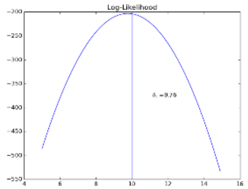
Association of Predicted Probabilities and Observed Responses			
Percent Concordant	100.0	Somers' D	1.000
Percent Discordant	0.0	Gamma	1.000
Percent Tied	0.0	Tau-a	0.571
Pairs	12	c	1.000



FIRTH MAXIMUM LIKELIHOOD, SEPARATION, 5 of 6

- **Solution: Use the “firth” option on the model statement.**
- proc logistic data = t descending;
- model y = x1 x2 / firth; run;

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X1	1.676	0.875	3.211
X2	1.612	0.467	5.567



FIRTH MAXIMUM LIKELIHOOD, SEPARATION, 6 of 6

- **Solution: Use the “firth” option on the model statement.**

Analysis of Penalized Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.9115	2.0557	2.0060	0.1567
X1	1	0.4972	0.4115	1.4598	0.2270
X2	1	0.3947	0.5944	0.4408	0.5067

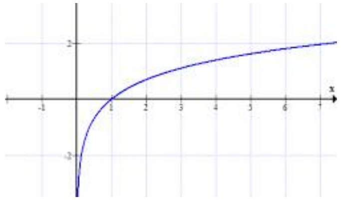
Association of Predicted Probabilities and Observed Responses			
Percent Concordant	100.0	Somers' D	1.000
Percent Discordant	0.0	Gamma	1.000
Percent Tied	0.0	Tau-a	0.571
Pairs	12	c	1.000

Logistic Regression Output (N = 116), Original Dataset

		Classical, Uadjusted		Classical, Adjusted	
Predictor of Recurrence	Ref	OR (95% CI)	OR P-Value	AOR (95% CI)	AOR P-Value
Age (years)	N/A	1.00 (0.96, 1.05)	0.907	1.00 (0.95, 1.05)	0.968
Female	Male	1.35 (0.32, 5.73)	0.689	1.11 (0.23, 5.27)	0.900
Cutaneous with sub-cutaneous extension sub-type	Cutaneous only	3.63 (0.86, 15.35)	0.080	4.49 (0.89, 22.74)	0.070
Extremity Location	Non-extremity	0.97 (0.23, 4.12)	0.969	1.07 (0.23, 4.97)	0.929
Grade 2 or 3	Grade 1	2.25 (0.48, 10.64)	0.720	1.86 (0.37, 9.20)	0.448
Unknown Grade	Grade 1	3.03 (0.46, 19.97)	0.413	4.89 (0.62, 38.23)	0.131

Logistic Regression Output (N = 115), Revised Dataset

		Classical, Uadjusted		Classical, Adjusted	
Predictor of Recurrence	Ref	OR (95% CI)	OR P-Value	AOR (95% CI)	AOR P-Value
Age (years)	N/A	1.00 (0.95, 1.04)	0.885	0.99 (0.94, 1.04)	0.569
Female	Male	0.90 (0.17, 4.70)	0.897	0.67 (0.13, 3.41)	0.630
Cutaneous with sub-cutaneous extension sub-type	Cutaneous only	**12.71 (1.51, 107.15)**	**0.019**	**17.24 (1.98, 150.09)**	**0.010**
Extremity Location	Non-extremity	1.46 (0.28, 7.59)	0.654	2.08 (0.41, 10.56)	0.376
Grade 2 or 3	Grade 1	3.37 (0.59, 19.36)	0.537	2.19 (0.44, 10.82)	0.337
Unknown Grade	Grade 1	4.54 (0.58, 35.24)	0.304	14.98 (1.38, 162.49)	0.026

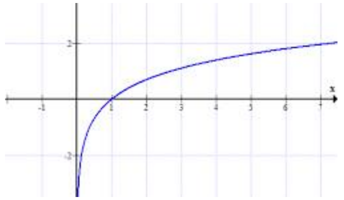


When Firth Maximum Likelihood Was Used, 1 of 2

- For both original and revised datasets, SAS did not give a warning in the log, “Maximum likelihood estimate may not exist.
- Did not use the Firth option for the original N = 116 dataset because there was no warning in the log and because the odds ratios were in a reasonable range.

- When the N = 115 dataset was analyzed with Proc Logistic, again, SAS did not give a warning about separation, although the confidence interval for the odds ratio for recurrence, based on sub-type, alone was large, 12.71 (1.51, 107.15). So, my first inclination was to use the Firth option.

- However, odds ratio calculation by hand was exactly what Proc Logistic produced without the Firth option. So, I only used the Firth option for the multi-variable logistic regression, because the original upper confidence bound was >999.



When Firth Maximum Likelihood Was Used, 2 of 2

Sub-Type	Recurred	Did not recur
Cutaneous	1	69
Sub-Cutaneous	7	38

- For univariable logistic regression, the odds ratio for recurrence by sub-type can be calculated by hand.
- Odds Ratio = (7/38)/(1/69) = 12.71, same as Proc Logistic without Firth.

- For multi-variable model, original odds ratio for sub-type = 45.85 (2.03, >999.999).
- By using “model recurrence = Age Female primary_subtype3 UpperLowerSite GradeCat3 / Firth”

- ➔Odds Ratio = 17.24 (1.98, 150.09).

Introduction to Bayesian Analysis (1 of 7)

- Bayes Theorem developed by Thomas Bayes, a Presbyterian minister and statistician, in the 1700's.

- Let A and B be events.
- P(A) = Initial belief or the probability of an event happening before you see any new evidence.
- P(B|A) = Probability of event B, given that event A has occurred.
- P(B) = Probability of event B, regardless of whether event A occurred.
- Posterior Probability P(A|B) = Updated probability of A, given that B has occurred.
- Bayes Theorem: P(A|B) = [P(B|A) * P(A)] / P(B).

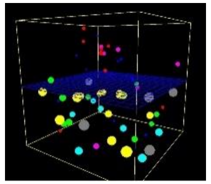
Non-Mathematical Bayesian Example (2 of 7)

- For a non-mathematical example, suppose two women, Mary and Jill sing side by side in a choir.

- Mary forms a prior belief about Jill’s personality, event A.

- Then, Mary and Jill travel to choir festival together, share a hotel room, and sing in the festival choir together, event B.

- After the conference, Mary’s view of Jill’s personality becomes A | B.

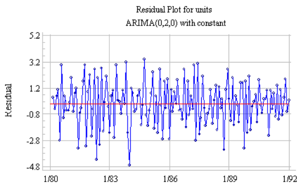


Bayesian Framework and Intervals (3 of 7)

- The Bayesian analysis framework is based on a prior distribution of data, that is updated by an experiment. Then, the posterior distribution is estimated after the experiment.
- The prior and posterior distributions are simulated based on the Markov Chain Monte Carlo (MCMC) method, which is a way of generating a large number of samples that will become uncorrelated with each other.

- The output of Bayesian regressions contains 95% credible intervals for the parameters, without the p-values and 95% confidence intervals in traditional frequentist statistics.

- The regression output looks similar to classical regression, but there are no p-values nor reliance on the asymptotic distributions of test statistics as N ➔∞ .
- Bayesian model goodness of fit are evaluated through diagnostic graphics and fit statistics.



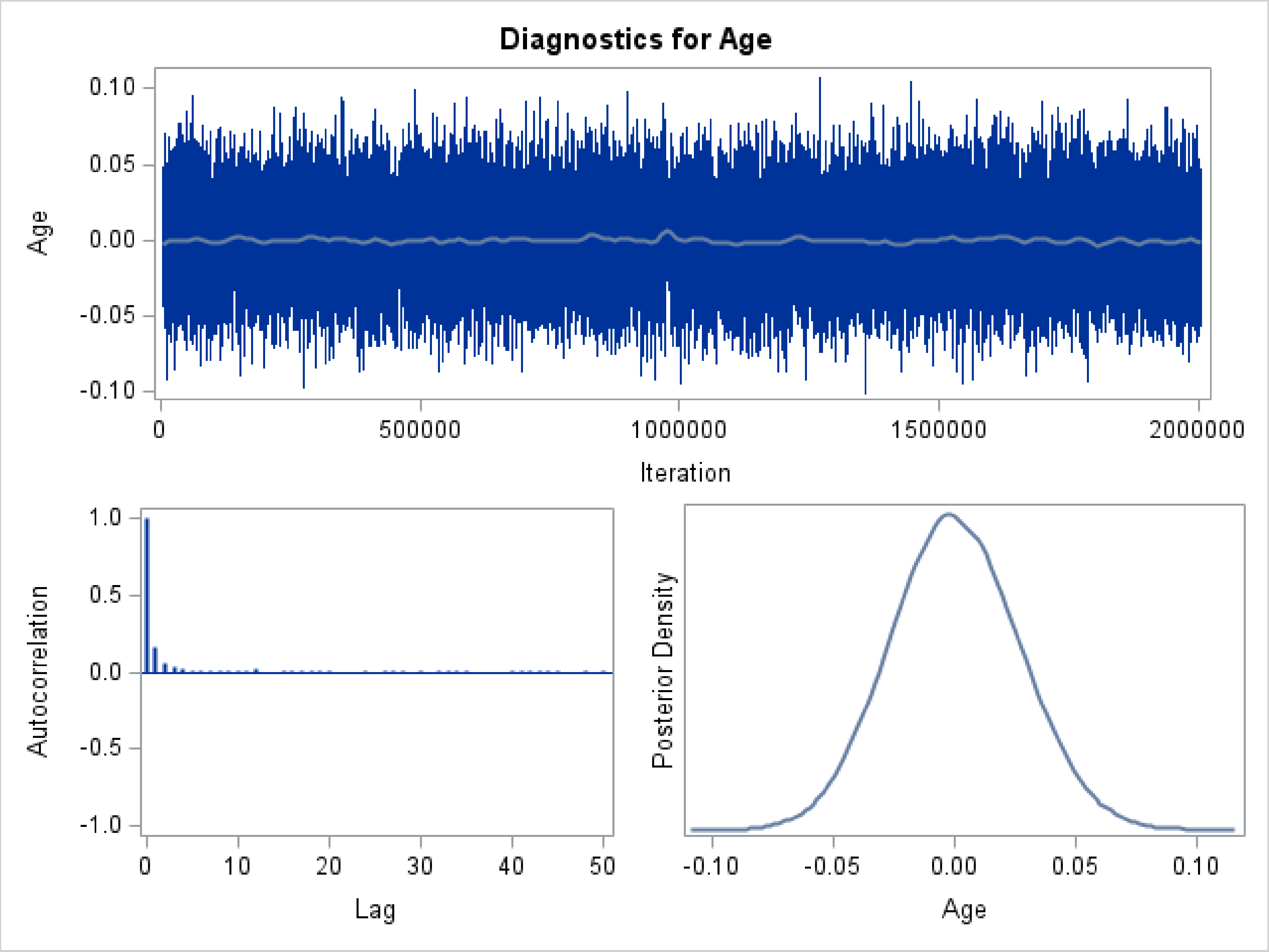
Bayesian Analysis (Diagnostic Graphs) (4 of 7)

- Graphically, Bayesian regression goodness of fit is evaluated by the trace plot, the auto-correlation plot, and the posterior density plot for each variable in the model.

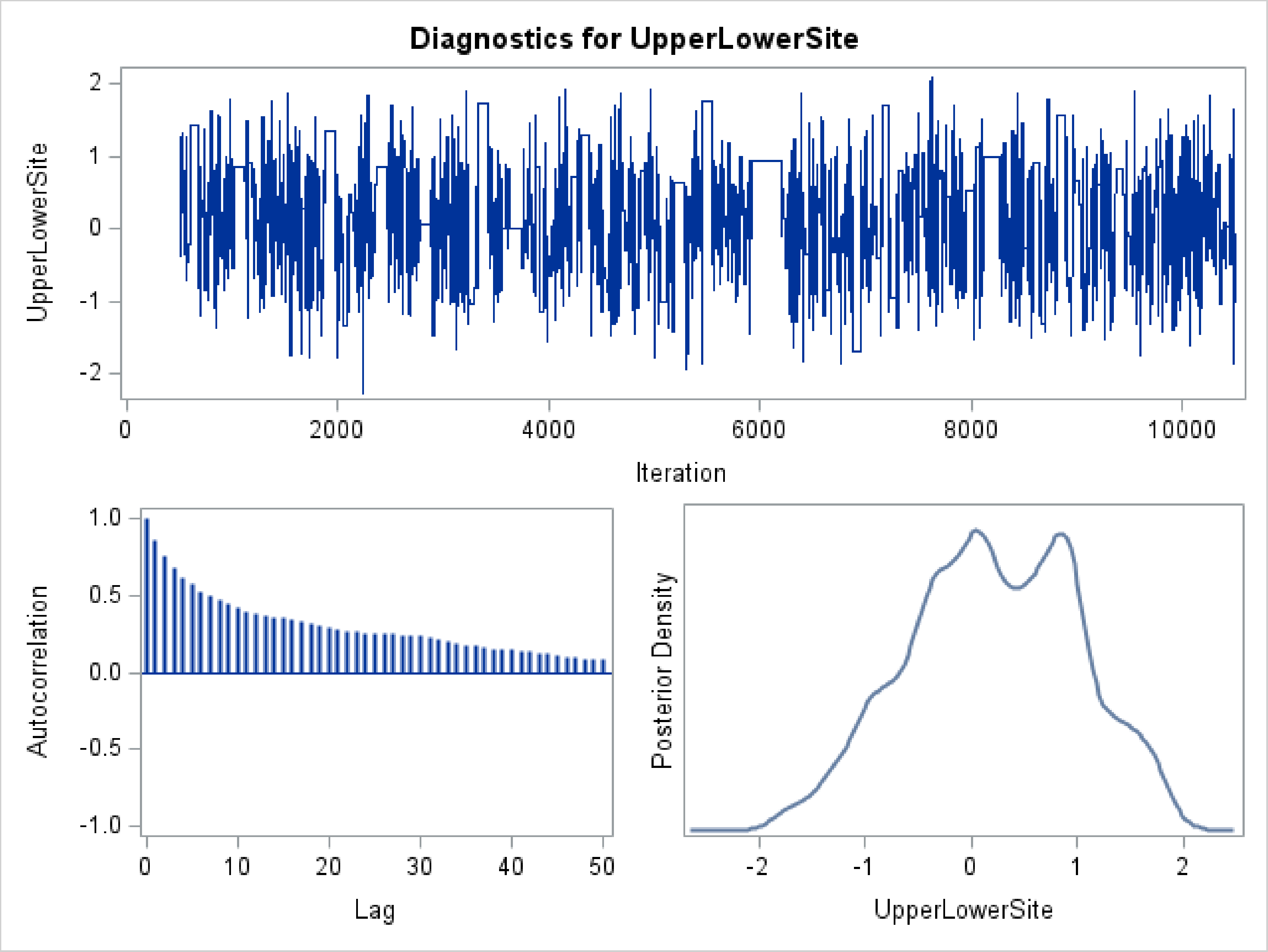
- A **trace plot** is a visual diagnostic tool used to assess the convergence of the Markov Chain Monte Carlo (MCMC) simulation. It plots the values of model parameters against the iteration number and should look like a caterpillar or white noise (no clear pattern).
- **Autocorrelation** refers specifically to dependencies between successive samples drawn from the posterior distribution using MCMC algorithms. High autocorrelation implies that the chain is slow in exploring the parameter space, which can lead to unreliable inference. The auto-correlation time should be less than 3 lags between successive samples.

- The **posterior density** plot is a histogram of posterior estimates. If the prior is normal, the posterior should be normal, or at least symmetrical.

Bayesian Analysis (Diagnostic Graphs for Good Model Fit) (5 of 7)



Bayesian Analysis (Diagnostic Graphs for Poor Model Fit) (6 of 7)



Bayesian Diagnostic Statistics (7 of 7)

Primary:

- The Geweke statistic tests whether the parameter means are approximately equal at the beginning and end of the Markov chain. Want a non-significant p-value.
- The proportion of the variance due to MCMC should be < 2.5%.
- The auto-correlation time should be < 3 lags.

- Other Diagnostics:
- Heidelberger-Welch stationarity test.
- Heidelberger-Welch halfwidth test: reports whether the sample size is adequate to meet the required accuracy for the mean estimate.
- Raftery-Lewis: evaluates the accuracy of the estimated percentiles.
- Gelman-Rubin: tests whether multiple chains would converge to the same distribution.

My Original Model (N = 116)

```
ods graphics on;
proc bglimm data=lms_jessie2025 nbi=5000 nmc=2000000 thin=40  seed=3172025
plots(smooth)=all diag=all stats=all NOCLPrint;
/* Burn-in replications, NBI = 5000 recommended, SAS default is 500 */
/* Increasing thinning parameter can reduce auto-correlation. */
/* I chose NMC to keep NMC/THIN ≥ 50000, may have to go larger. */

class  primary_subtype3 (ref='0’);
model recurrence(descending)  = Age Female primary_subtype3 / dist=binary link=logit
coeffprior=normal(var=1e6);
/* Minimally informative prior Normal(mean 0, variance 1,000,000 */

Estimate 'Cutaneous Recurrence' int 1 Age 59.5 Female .270  primary_subtype3 0 1
UpperLowerSite .678  GradeCat3 .34 .13 .53;
Estimate 'Sub-Cutaneous Recurrence' int 1 Age 59.5 Female .270  primary_subtype3 1 0
UpperLowerSite .678  GradeCat3 .34 .13 .53;
run;  ods graphics off;
```

My Original Model Output (N = 116)

		Classical, Adjusted		Bayesian	
Predictor of Recurrence	**Ref**	**AOR (95% CI)**	**AOR P-Value**	**AOR, Equal Tails**	**AOR, Highest Posterior Density**
Age (years)	N/A	1.00 (0.95, 1.05)	0.968	1.00 (0.95, 1.05)	1.00 (0.95, 1.05)
Female	Male	1.11 (0.23, 5.27)	0.900	1.04 (0.20, 5.03)	1.04 (0.21, 5.17)
Cutaneous with sub-cutaneous extension sub-type	Cutaneous only	4.49 (0.89, 22.74)	0.070	**5.48 (1.10, 29.74)**	**5.48 (1.06, 28.20)**
Extremity Location	Non-extremity	1.07 (0.23, 4.97)	0.929	1.16 (0.25, 5.91)	1.16 (0.25, 5.81)
Grade 2 or 3	Grade 1	1.86 (0.37, 9.20)	0.448	1.97 (0.39, 10.06)	1.97 (0.38, 9.70)

Model Output: Logistic Regression Credible Intervals Before Exponentiation (N = 116)

Posterior Intervals					
Parameter	**Alpha**	**Equal-Tail Interval**		**HPD Interval**	
Age	0.05	-0.051	0.054	-0.052	0.052
Female	0.05	-1.599	1.615	-1.56	1.644
primary_subtype3	0.05	0.1	3.392	0.059	3.339
UpperLowerSite	0.05	-1.381	1.777	-1.391	1.759
GradeCat3	0.05	-0.939	2.309	-0.967	2.272

- ✓ **Note that Primary Subtype credible interval is > 0.**

Model Output: Auto-Correlation Time (N = 116)

Effective Sample Sizes	
Parameter	Autocorrelation Time
Intercept	1.7607
Age	1.6102
Female	1.5533
primary_subtype3	1.5916
UpperLowerSite	1.6577
GradeCat3	1.5395

✓ **Note that all auto-correlation times < 3 lags.**

Model Output: Monte Carlo Standard Errors (N = 116)

Monte Carlo Standard Errors			
Parameter	**MCSE**	**Standard Deviation**	**MCSE/SD**
Intercept	0.0103	1.7372	0.00593
Age	0.000151	0.0266	0.00567
Female	0.00456	0.8183	0.00557
primary_subtype3	0.00475	0.8424	0.00564
UpperLowerSite	0.00466	0.8087	0.00576
GradeCat3 2	0.00463	0.8344	0.00555

- ✓ **Note that all MCSE/SD values < .025 (2.5%).**

Model Output: Monte Carlo Standard Errors (N = 116)

Monte Carlo Standard Errors			
Parameter	MCSE	Standard Deviation	MCSE/SD
Intercept	0.0103	1.7372	0.00593
Age	0.000151	0.0266	0.00567
Female	0.00456	0.8183	0.00557
primary_subtype3	0.00475	0.8424	0.00564
UpperLowerSite	0.00466	0.8087	0.00576
GradeCat3 2	0.00463	0.8344	0.00555

✓ **Note that all MCSE/SD values < .025 (2.5%).**

Model Output: Geweke Tests (N = 116)

Geweke Diagnostics		
Parameter	**z**	**Pr > |z|**
Intercept	-0.5779	0.5634
Age	0.4273	0.6692
Female	-0.2885	0.773
primary_subtype3	-0.6165	0.5376
UpperLowerSite	0.1001	0.9202
GradeCat3 2	0.838	0.402

✓ **Note that all Geweke test p-values are non-significant (p > .05)**

Model Output: Heidelberger-Welch Diagnostics (N = 116)

Heidelberger-Welch Diagnostics							
Parameter	Stationarity Test			Half-Width Test			
	Cramer-von Mises Stat	p-Value	Test Outcome	Half-Width	Mean	Relative Half-Width	Test Outcome
Intercept	0.0878	0.6481	Passed	0.0202	-4.3291	-0.0047	Passed
Age	0.1997	0.2681	Passed	0.00034	0.00016	2.1042	Passed
Female	0.1475	0.3973	Passed	0.00823	0.0432	0.1903	Passed
primary_subtype3 1	0.1	0.5848	Passed	0.0101	1.7009	0.0059	Passed
UpperLowerSite	0.0994	0.5878	Passed	0.0132	0.1501	0.0882	Passed
GradeCat3 2	0.1231	0.483	Passed	0.00915	0.6777	0.0135	Passed
GradeCat3 4	0.4237	0.0626	Passed	0.0259	1.6847	0.0154	Passed

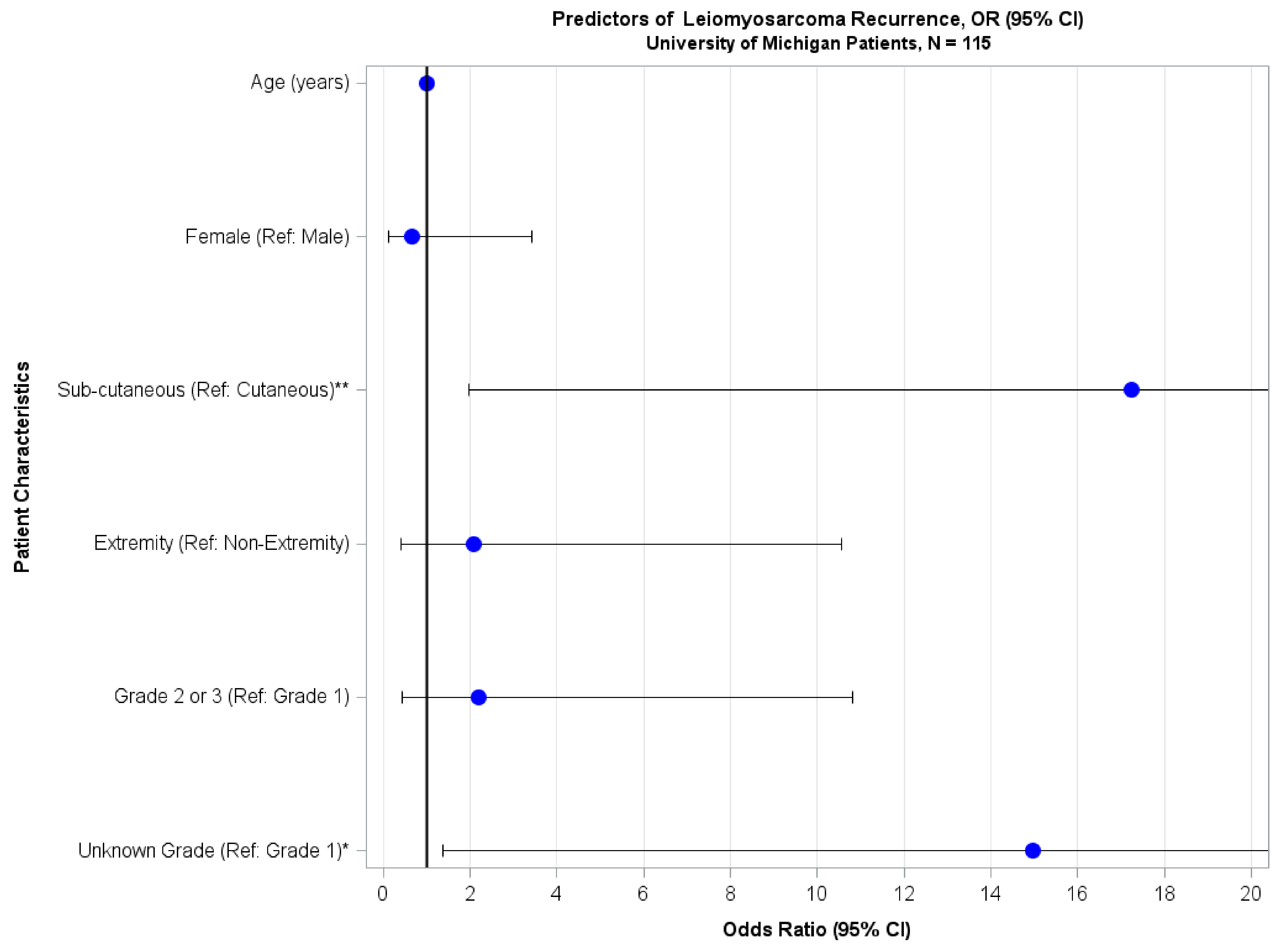
- ✓ **Note that all Heidelberger-Welch tests listed a Passed or Failed.**

My Modified Model (N = 115)

```
/* Proc GenMod worked better with modified dataset; Proc BGLIMM didn’t converge */

ods graphics on;
proc genmod data = lms_jessie2025 descending;
  class  primary_subtype3 (ref='0') GradeCat3(ref='1');
  model recurrence  = Age Female primary_subtype3 UpperLowerSite GradeCat3 /link=logit
dist=bin;

/* Bayes statement in Proc GenMod */
bayes seed=3172025 coeffprior=normal plots=all nbi=5000 nmc=750000 thin=75 diag=all
stats=all;

run;
ods graphics off;
```

My Modified Model Output (N = 115)

		Classical, Adjusted		Bayesian	
Predictor of Recurrence	**Ref**	**AOR (95% CI)**	**AOR P-Value**	**AOR, Equal Tails**	**AOR, Highest Posterior Density**
Age (years)	N/A	0.99 (0.94, 1.04)	0.569	0.98 (0.92,1.05)	0.98 (0.91,1.05)
Female	Male	0.67 (0.13, 3.41)	0.630	0.40 (0.04,3.14)	0.40 (0.04,3.42)
Cutaneous with sub-cutaneous extension sub-type	Cutaneous only	**17.24 (1.98, 150.09)**	**0.010**	177.08 (6.7,23,890)	177.08 (4.5,11,535)
Extremity Location	Non-extremity	2.08 (0.41, 10.56)	0.376	3.98 (0.53,46.37)	3.98 (0.48,39.91)
Grade 2 or 3	Grade 1	2.19 (0.44, 10.82)	0.337	3.13 (0.45,31.01)	3.13 (0.40,26.71)



Predictors of Leiomyosarcoma Recurrence, OR (95% CI)
University of Michigan Patients, N = 115
Patient Characteristics
Age (years)
Female (Ref: Male)
Sub-cutaneous (Ref: Cutaneous)**
Extremity (Ref: Non-Extremity)
Grade 2 or 3 (Ref: Grade 1)
Unknown Grade (Ref: Grade 1)*
0
2
4
6
8
10
12
14
16
18
20
Odds Ratio (95% CI)

SAS Code for Odds Ratio Plot (1 of 2)

```
/* Odds Ratio Plot with 95% CI, Firth Adjustment */
Data CIOR;
infile datalines dlm='#';
Length Outcome $32;
Format Outcome $char32. Est LCL UCL 7.2 Ord 3.0;
Input Ord Outcome $ Est LCL UCL;
Datalines;
1 # Age (years) # 0.99 # 0.94 # 1.04
2 # Female (Ref: Male) # 0.67 # 0.13 # 3.41
3 # Sub-cutaneous (Ref: Cutaneous)** # 17.24 # 1.98 # 150.09
4 # Extremity (Ref: Non-Extremity) # 2.08 # 0.41 # 10.56
5 # Grade 2 or 3 (Ref: Grade 1) # 2.19 # 0.44 # 10.82
6 # Unknown Grade (Ref: Grade 1)* # 14.98 # 1.38 # 162.49
;
Run;
```

SAS Code for Odds Ratio Plot (2 of 2)

```
ods graphics on / width=11 in;
Title 'Predictors of  Leiomyosarcoma Recurrence, OR (95% CI)' font="Arial" height=14pt;;
Title2 'University of Michigan Patients, N = 115' font="Arial" height=12pt;
Proc SGPlot data=CIOR;
Scatter X=Est Y=Outcome / XerrorLower=LCL XErrorUpper=UCL
MarkerAttrs=OR(symbol=CircleFilled Size=12 color=blue)
ERRORBARATTRS=(color=black);
Refline 1 / Axis=X LINEATTRS=(thickness=2 color=black) ;

xaxis label='Odds Ratio (95% CI)' LABELATTRS=(Size=11 Weight=Bold)
VALUEATTRS=(Size=11) values=(0 to 20 by 2) grid;
yaxis label='Patient Characteristics' LABELATTRS=(Size=11 Weight=Bold)
VALUEATTRS=(Size=11) DISCRETEORDER=DATA;
Run;
ods graphics off;
```

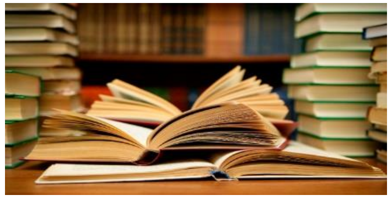
Conclusions

- The sarcoma publication will use the results from the modified dataset with N = 115 patients.

- The logistic regression results from Proc Logistic, with the Firth option, will be reported for the main outcomes in the paper.

- Initially, Bayesian analysis with Proc BGLIMM looked like the best option. Due to quasi-separation occurring after the dataset was modified, the analysis changed. Classical regression, with the Firth option, gave the most stable and believable results. The Bayesian logistic regression generated huge confidence intervals.

- Although Bayesian regression is sometimes a good alternative to classical regression for small datasets, classical logistic regression produced better results in this case. The Bayesian logistic regression was reported in an appendix to further confirm the results from the classical logistic regression and show that we left no stone unturned.



References

- Allison, Paul. Logistic Regression Using SAS: Theory and Application, 2nd ed (2012). SAS Institute Publications.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer-Verlag.
- Firth, D. (1993). "Bias Reduction of Maximum Likelihood Estimates." Biometrika 80:27–38.
- Geweke, J. (1992). "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments." In *Bayesian Statistics, vol. 4*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 169–193. Oxford: Clarendon Press.
- Heinze, G., and Schemper, M. (2002). "A Solution to the Problem of Separation in Logistic Regression." Statistics in Medicine 21:2409–2419.

Contact Information

- Brandy R. Sinco, Statistician and Programmer/Analyst
- Michigan Medicine
- CHOP (Center for Healthcare Outcomes and Policy)
- Ann Arbor, MI 48109-2800
- E-Mail: brsinco@umich.edu