Paper JP-01

JMP® Analytics Applied in Diagnostic Radiology and Neurosurgery Trauma Research

Melvin Alexander, University of Maryland Medical Center, Baltimore, MD

ABSTRACT

JMP® analytics helped trauma neurosurgeons and radiologists, diagnosing brain and spinal-cord injured patients, identify key factors affecting patient outcomes. A study in *Neurosurgical Focus* (June, 2009, 26, E8) showed how JMP discovered risk factors associated with subdural hygroma (SDG). SDG is the collection of cerebrospinal fluids over the brain following surgery that can result in delayed head-trauma complications. Sixty-eight patients were followed 36-weeks longitudinally, 39 patients had SDG, 29 patients without SDG. Likelihood Ratio and Fisher's exact test results indicated that motor vehicle accidents ($p< 0.007$) and falls ($p<0.005$) were most often linked with SDG development. Diffuse brain injuries were more prone to SDG complications ($p< 0.0299$).

JMP was used in another study published in the *American Journal of Roentgenology* (2009, 192, 52-58) to identify quantitative, anatomic measurements of head and neck images that distinguished patients with and without Craniocervical Distraction injuries (CCDIs). CCDIs are often fatal head-neck injuries that have been associated with survival to the hospital. Improved emergent patient retrieval systems have increased survival to the hospital. However, CCDIs tend to be missed in physical examination and diagnostic imaging. Logistic regression and recursive partitioning determined that measures such as the midline occiput-C1 spinolaminar distance ($p=0.0016$), midline C1-C2 spinolaminar distance ($p<0.0001$), basion-dens distance ($p<0.0001$), sum of condylar displacement ($p=0.0002$), and basion-posterior axial line distance ($p<0.0001$) differentiated patients with CCDIs from patients without CCDIs.

This presentation examines the role JMP® software (version 7 or higher) plays in advancing follow-on investigations.

INTRODUCTION

This presentation will show how JMP® analytics assists trauma neurosurgeons and radiologists identify key factors affecting patient outcomes from reviews of CT Imaging.

As a statistical consultant I assist researchers to:

- plan investigations, e.g., determine power and sample size calculations;
- write statistical methods sections for Institutional Review Board (IRB) submissions and grant proposals;
- design experimental plans for data collection and integrity protocols for data safety monitoring, etc.;
- provide data analyses, data visualizations, and fit statistical models that describe data and establish relationships between factors on the response variables of interest;
- predict and simulate effects of factors on patient outcomes that help answer questions posed by researchers;
- write articles for medical journals, prepare posters and presentations for medical conferences.

JMP® plays an important role in data extraction, manipulation, assembly, importing images and data from other databases. Scripts allow easy ways of documenting and reproducing analyses. JMP is an important tool that helps me perform these tasks in a timely manner.

Two case studies are featured: (1) a study of risk factors associated with the build-up of cerebrospinal fluids over the brain after delicate surgical procedure for treating severe traumatic brain injuries (STBIs); (2) a study to identify factors that distinguished patients with and without a special type of head-neck injury that was often missed during CT imaging.

Finally, a discussion follows about the increased role JMP software plays in advancing follow-on studies.

CASE STUDY 1

The first case study is taken from Aarabi et al. [1]. The objective was to identify which risk factors had significant influences on the follow-up outcomes of severe traumatic brain injury (STBI) patients developing subdural hygroma (SDG) following decompressive crainiectomy (DC). Severe traumatic brain injury remains a significant cause of mortality and morbidity worldwide. Increased intracranial pressure (brain swelling), occurring after a severe traumatic brain injury event, is a common and potentially devastating phenomenon.

To understand the medical aspects of this case study, some key terms are defined as follows:

Intracranial Pressure (ICP): Increased pressure inside the skull which can result in coma, neurological complications, or death for patients with STBIs that occurred from events such as motor vehicle crashes, falls, or sports injuries.

Decompressive Craniectomy (DC): a neurosurgical procedure that removes part of the skull to allow the brain room to expand without being squeezed and relieve ICP. DC is considered a last-resort surgical procedure for reducing ICP when patients are unresponsive to other medical treatments.

Cerebrospinal Fluids (CSFs): water-like fluids that surround tissues, cells, the brain, and spinal cord. CSF functions are not well understood but are believed to support and protect the brain and spinal cord.

Subdural Hygroma (SDG): Collection of CSFs over the brain and spinal cord that may result in delayed head-trauma complications. SDG is usually observed in the first week after surgery . CSF build-up tends to increase up to one month. Excess SDG may require surgical drainage to avoid neurological deterioration if left untreated.

Prior to this study the factors contributing to the development of SDG were not clearly understood. Researchers were interested in investigating the relationship between risk factors and SDG after DC.

Figure 1 shows a schematic representation of the interaction between the cellular layers, membranes, tissues, and CSFs as observed from an electron microscope

Figure 1. Schematic Representation of dural border cells as observed by an electron microscope (Reprinted with permission from *Neurosurgical Focus*)

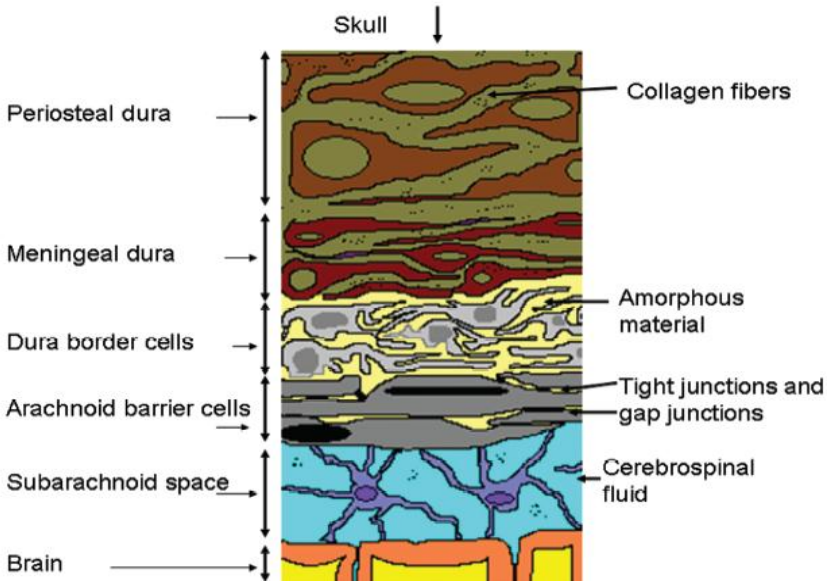


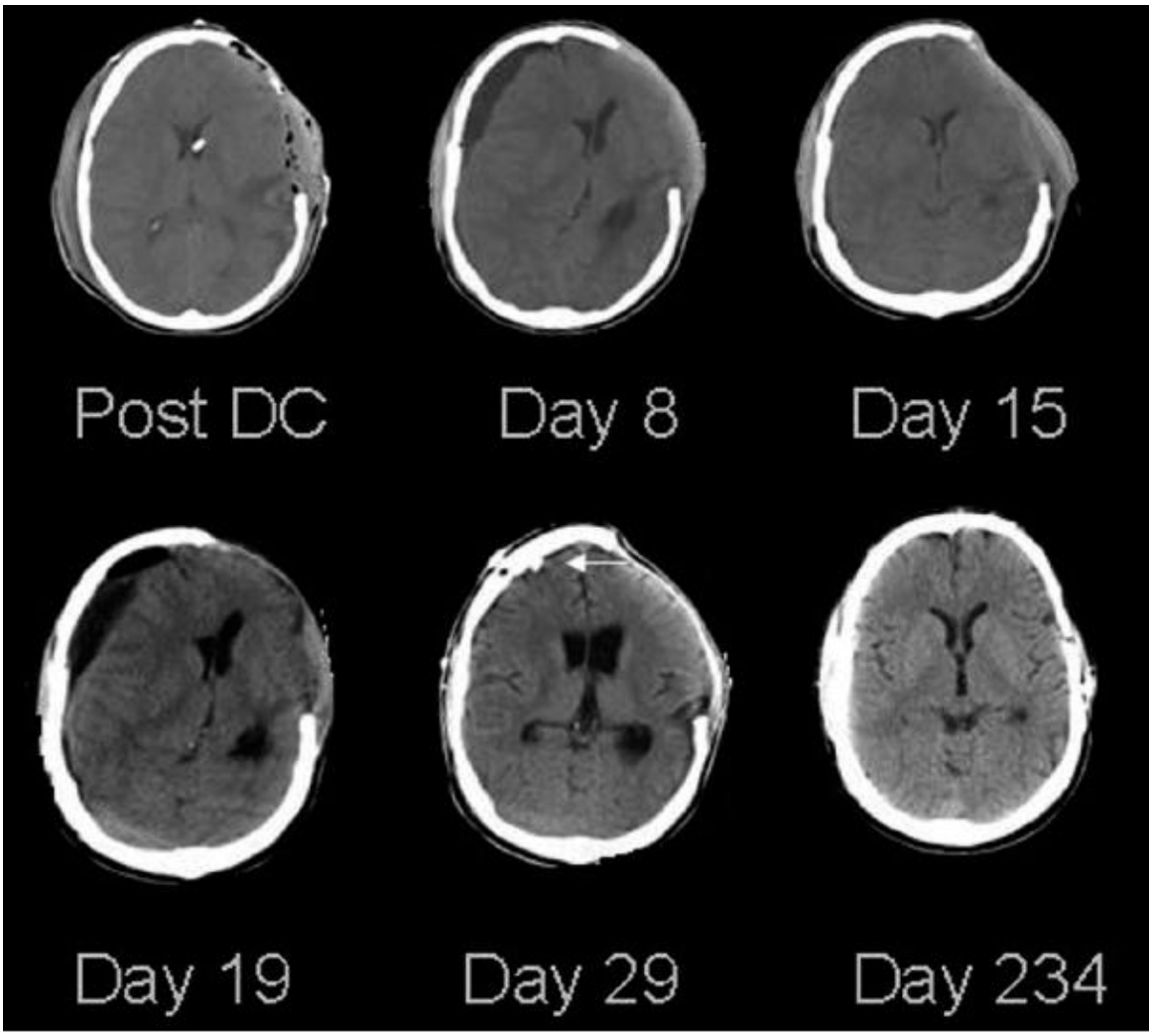
Figure 1 indicates a lateral view of the development and evolution of SDG. The bottom part of Figure 1 shows the layers of cells and CSFs that form around the brain.

METHODS

We followed 68 patients who survived DC longitudinally over 36 weeks after surgery. CT imaging scans were retrieved from 39 patients who received SDG and 29 patients without SDG. We studied the clinical and imaging characteristics of both groups of patients to understand the dynamics of SDG appearances, changes in CSF volumes, and complications of patients.

Likelihood ratio, Fisher's exact tests and odds ratio calculations were fitted using JMP (versions 7 and higher, SAS Institute Inc., Cary, NC, 1989-2009) data table, analysis platforms, and scripts.

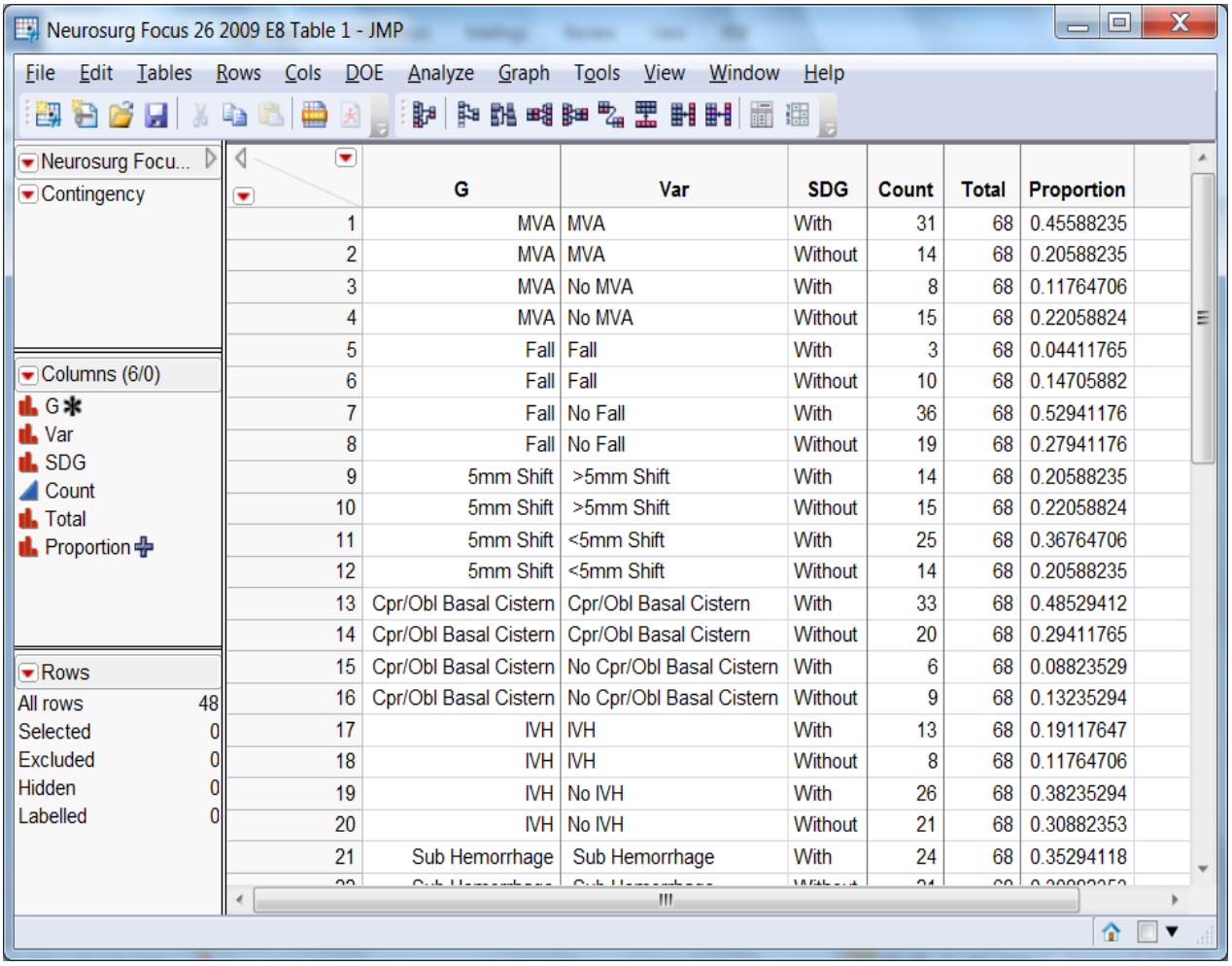
Figure 2. CT scans of a 24-year old male who underwent DC after being involved in a motor vehicle accident (Reprinted with permission from *Neurosurgical Focus*)



The top left scan of Figure 2 (Post DC) showed CSF build-up after surgery. On Day 8, a second CT scan on the patient showed signs of mental decline for the patient . The midline structures shifted to the left on Day 19. On Day 29, the volume of CSFs was excessive and required drainage by insertion of a twist drill bur hole. The patient remained neurologically intact for 47 months after discharge (Day 234) from the medical center.

Table 1 is a subset of the JMP data table of risk factors and SDG accumulation in 68 patients.

Table 1. Portions of the JMP SDG Post DC Data Table of 68 patients

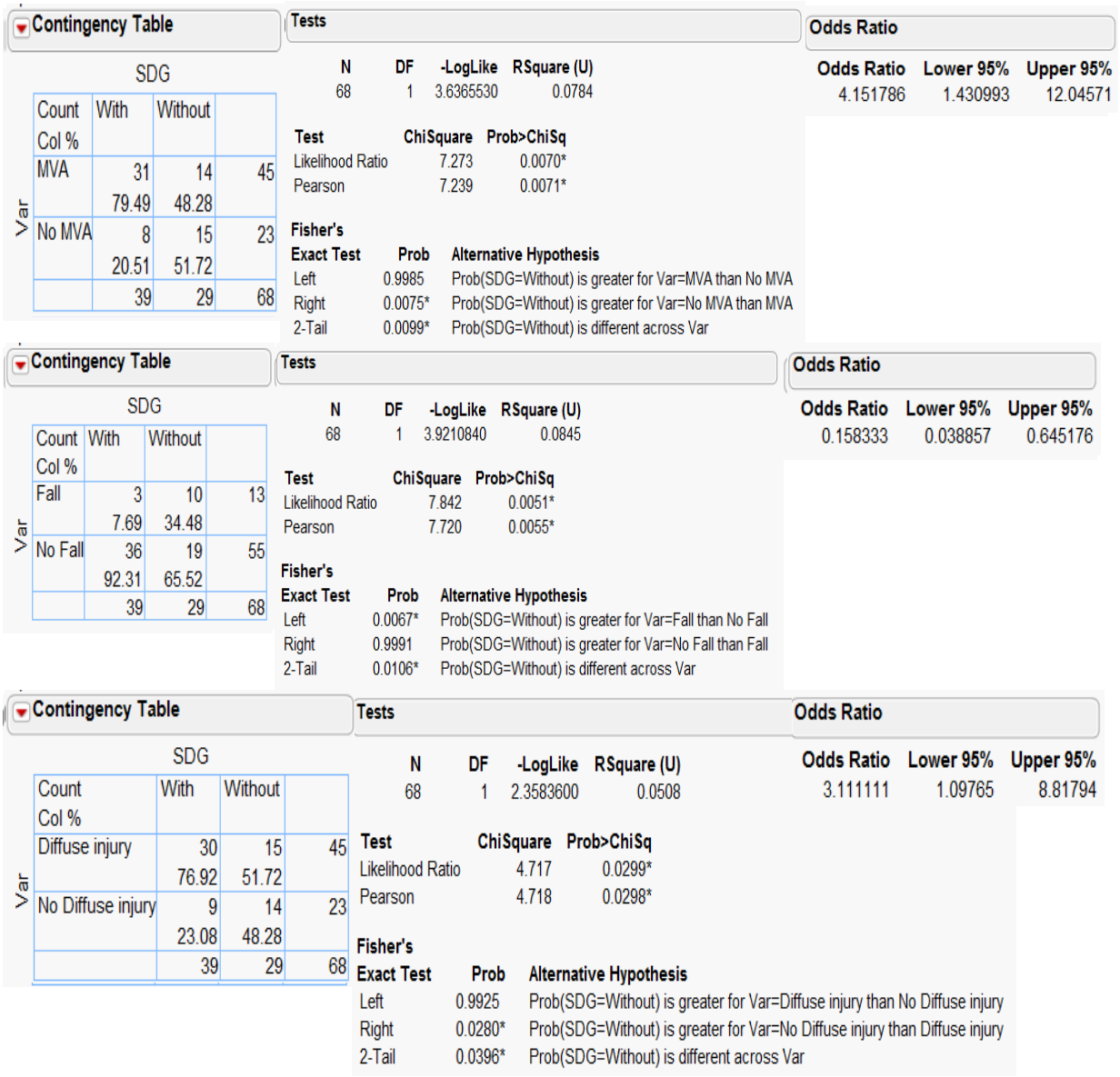


	G	Var	SDG	Count	Total	Proportion
1	MVA	MVA	With	31	68	0.45588235
2	MVA	MVA	Without	14	68	0.20588235
3	MVA	No MVA	With	8	68	0.11764706
4	MVA	No MVA	Without	15	68	0.22058824
5	Fall	Fall	With	3	68	0.04411765
6	Fall	Fall	Without	10	68	0.14705882
7	Fall	No Fall	With	36	68	0.52941176
8	Fall	No Fall	Without	19	68	0.27941176
9	5mm Shift	>5mm Shift	With	14	68	0.20588235
10	5mm Shift	>5mm Shift	Without	15	68	0.22058824
11	5mm Shift	<5mm Shift	With	25	68	0.36764706
12	5mm Shift	<5mm Shift	Without	14	68	0.20588235
13	Cpr/Obl Basal Cistern	Cpr/Obl Basal Cistern	With	33	68	0.48529412
14	Cpr/Obl Basal Cistern	Cpr/Obl Basal Cistern	Without	20	68	0.29411765
15	Cpr/Obl Basal Cistern	No Cpr/Obl Basal Cistern	With	6	68	0.08823529
16	Cpr/Obl Basal Cistern	No Cpr/Obl Basal Cistern	Without	9	68	0.13235294
17	IVH	IVH	With	13	68	0.19117647
18	IVH	IVH	Without	8	68	0.11764706
19	IVH	No IVH	With	26	68	0.38235294
20	IVH	No IVH	Without	21	68	0.30882353
21	Sub Hemorrhage	Sub Hemorrhage	With	24	68	0.35294118

Column G stratified the different injury mechanisms and risk factors of interest. The Var column denoted the dichotomous states of patient characteristics and injury mechanisms (X rows variable) for the Contingency Table analysis. The SDG column (Y variable) described the presence or absence of the response outcomes in the 2 x 2 table using the Fit Y by X Analysis platform.

Appendix 1 lists the JMP script that produced Tables 1, 2 (Contingency Table Analyses), 3 (Odds Ratio Table), and Figure 3 (Odds Ratio and 95% Confidence Interval Plot).

Table 2. Contingency Table of Likelihood Ratio and Fisher's Exact Tests of Selected Significant Risk Factors



Contingency Table

Var	Count / Col %	With	Without	
	MVA	31	14	45
		79.49	48.28	
	No MVA	8	15	23
		20.51	51.72	
		39	29	68

Tests

N	DF	-LogLike	RSquare (U)
68	1	3.6365530	0.0784

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	7.273	0.0070*
Pearson	7.239	0.0071*

Fisher's

Exact Test	Prob	Alternative Hypothesis
Left	0.9985	Prob(SDG=Without) is greater for Var=MVA than No MVA
Right	0.0075*	Prob(SDG=Without) is greater for Var=No MVA than MVA
2-Tail	0.0099*	Prob(SDG=Without) is different across Var

Odds Ratio

Odds Ratio	Lower 95%	Upper 95%
4.151786	1.430993	12.04571

Contingency Table

Var	Count / Col %	With	Without	
	Fall	3	10	13
		7.69	34.48	
	No Fall	36	19	55
		92.31	65.52	
		39	29	68

Tests

N	DF	-LogLike	RSquare (U)
68	1	3.9210840	0.0845

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	7.842	0.0051*
Pearson	7.720	0.0055*

Fisher's

Exact Test	Prob	Alternative Hypothesis
Left	0.0067*	Prob(SDG=Without) is greater for Var=Fall than No Fall
Right	0.9991	Prob(SDG=Without) is greater for Var=No Fall than Fall
2-Tail	0.0106*	Prob(SDG=Without) is different across Var

Odds Ratio

Odds Ratio	Lower 95%	Upper 95%
0.158333	0.038857	0.645176

Contingency Table

Var	Count / Col %	With	Without	
	Diffuse injury	30	15	45
		76.92	51.72	
	No Diffuse injury	9	14	23
		23.08	48.28	
		39	29	68

Tests

N	DF	-LogLike	RSquare (U)
68	1	2.3583600	0.0508

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	4.717	0.0299*
Pearson	4.718	0.0298*

Fisher's

Exact Test	Prob	Alternative Hypothesis
Left	0.9925	Prob(SDG=Without) is greater for Var=Diffuse injury than No Diffuse injury
Right	0.0280*	Prob(SDG=Without) is greater for Var=No Diffuse injury than Diffuse injury
2-Tail	0.0396*	Prob(SDG=Without) is different across Var

Odds Ratio

Odds Ratio	Lower 95%	Upper 95%
3.111111	1.09765	8.81794

Fisher's exact tests supplemented the Pearson and Likelihood Ratio tests that would apply when the two-way tables had one or more cells less than five or were highly imbalanced. The SAS® equivalent of Table 2 can be made using PROCs FREQ (with the CMH option in the TABLE statement), LOGISTIC (in the "Odds Ratio Estimates" table), GENMOD (with the ESTIMATE statement), and PHREQ (labeled as "Hazard Ratio" with the main effects RISKLIMITS option in the MODEL statement or interactions HAZARDRATIO statement).

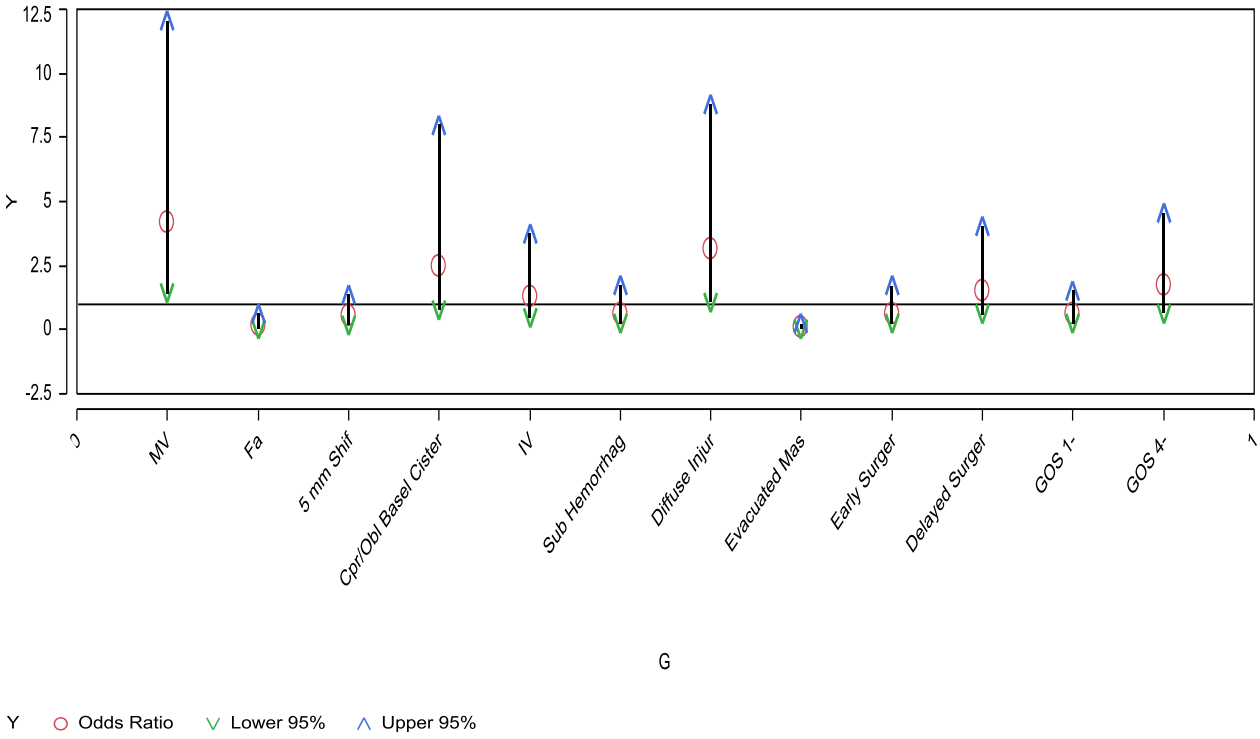
The Odds Ratios in Table 3 measured the association between the risk factors and having SDG. Table 3 can be extracted by placing the pointer in the Odds Ratio table of the Contingency Table report and selecting "Make Combined Data Table". Next add the G column using the **Table > Summarize** dialog. Finally, the two tables are joined together with the **Table > Join** dialog.

Step 2 in the JSL script of Appendix 1 may also be used to produce the Odds Ratio Data Table.

Table 3. Odds Ratio Tables and 95% Confidence Intervals

	G	Odds Ratio	Lower 95%	Upper 95%
1	MVA	4.1517857143	1.4309926107	12.045711828
2	Fall	0.1583333333	0.0388567477	0.645176087
3	5 mm Shift	0.5226666667	0.1963493913	1.3912976384
4	Cpr/Obl Basel Cistern	2.475	0.7659800276	7.9971079918
5	IVH	1.3125	0.458451844	3.7575511417
6	Sub Hemorrhage	0.6095238095	0.2156809541	1.7225409446
7	Diffuse Injury	3.1111111111	1.0976501005	8.8179396528
8	Evacuated Mass	0.0625	0.0184910583	0.2112507537
9	Early Surgery	0.6492753623	0.2465488921	1.7098373168
10	Delayed Surgery	1.5401785714	0.5848509622	4.0559906456
11	GOS 1-3	0.5833333333	0.2204788644	1.5433578124
12	GOS 4-5	1.7142857143	0.6479378871	4.5355821424

Figure 3. Overlay Plot of Odds Ratios and 95% Confidence Intervals of Risk Factors from Table 3



Significant Odds in Figure 3 showed that motor vehicle accidents (MVAs) and Diffuse Injuries (injuries spread over wider areas of the brain due to rapid acceleration or deceleration) had higher odds of having SDGs than patients without SDGs since the 95% confidence intervals did not include one. Falls and evacuated mass had significantly lower odds of having SDGs. Clinically, the odds ratios from MVAs and diffuse injuries tend to be higher because sheering stress produced by kinetic energy can tear arachnoid barrier cells. The effects tend to damage the integrity of the dura border cells and arachnoid interface layers (see Figure 1 middle).

CASE STUDY 2

The second case study came from Chang, Alexander, and Mirvis [2]. The goal of this study was to identify quantitative, anatomic measurements of head and neck images that would distinguish patients with and without Craniocervical Distraction injuries (CCDIs). Craniocervical Distraction injuries (CCDIs) are a class of injuries that involve the skull base, the atlas, and the axis. CCDIs are often fatal and have been associated with survival to the hospital. Until recently, improved emergent patient retrieval systems have increased survival to the hospital. However, CCDIs tend to be hard to identify and may be missed in physical examination and diagnostic imaging. With increased use of CT imaging, specific criteria that can identify normal and abnormal CT anatomy in the upper cervical spine are not well defined. Failure to identify patients with and without CCDIs could result in further complications that delay recovery.

CCDIs are typically characterized by abnormal extensions of the atlanto-occipital joints (i.e., joints that allow the head to nod up and down along the vertical spinal column).

METHODS

CT scans were collected from patients who had imaging for cervical collar application after they were admitted to the University of Maryland's R Adams Cowley Shock Trauma Center from 2000-2006. Certified technologists reformatted and measured the anatomic features of the images for patients with CCDIs (n=35) and without CCDIs (n=50) confirmed at discharge. Two independent radiologists, blind to patient outcomes, measured the CT scans of each patient using a Picture Archival and Communications System (PACS) workstation and electronic calipers.

Table 4 describes the measurements made on the reformatted images by the technologist and two independent radiologists.

Table 4. Name and Description of Measurements Made on CT Images

Measurements	Description
Midline Occiput-C1 spinolaminar line	Distance in mm from the opisthion to the top of the arch of the atlas as measured in the midline sagittal MPR
Basion-dens	Distance in mm from the basion to the top of the dens as measured in a midline sagittal MPR [5]
Condylar displacement (condylar sum)	Distance from anterior-posterior midpoints of the occipital condyle and fossa of the atlas measured in the sagittal MPR through the midpoint of the joint
Midline C1-C2 spinolaminar line	Distance in mm from the bottom of the arch of C1 to the top of the arch of C2 as measured in the midline sagittal MPR
Atlanto-dental interval (ADI)	Distance in mm from the posterior surface of the C1 anterior arch to the anterior surface of the dens in the midline sagittal MPR
V-sign	Qualitative evaluation of the angle made between the posterior surface of the C1 vertebral body and the anterior surface of the dens (convergent/parallel, divergent)
Vertical atlanto-dens line	Vertical distance measured in mm from the top of the C1 vertebral body to the top of the dens measured on the sagittal midline MPR
Posterior axial line (PAL)	Length of a perpendicular line drawn from the basion to a line tangential to the posterior cortical surface of the axis (PAL) [5]
C1-C2 distraction	Qualitative evaluation of the distance between the atlas and axis in the coronal MPR plane at level of the root of the transverse process
ST edema (mid C1)	Thickness in mm of the soft tissues anterior to the C2 vertebral body in the midline sagittal projection

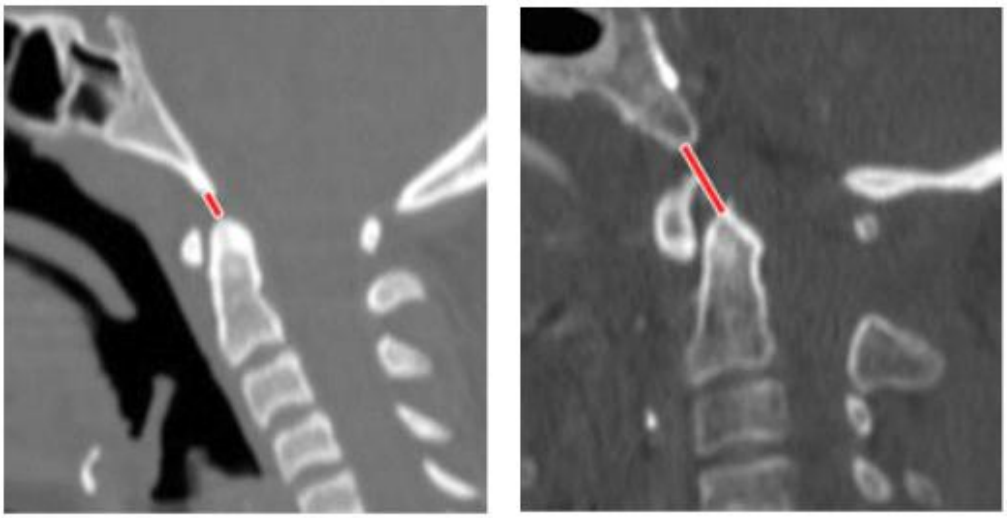
The script below produced Figures 4a and 4b and demonstrated the image processing functionality that could be used to import CT images from the PACS system into the JMP's Graphic Box for viewing by radiologists.

```
fig1a = NewImage("J:\MAJUG\Feb 2009\AJR192_Fig1a.gif");
w = New Window( "Figure 1A (Reprinted with permission from the American Journal of
Roentgenology)
",
   gb = Graph Box(
      Frame Size( 250, 250 ),X Scale( 0, 80 ), Y Scale( 0, 80 ),
      << Add Image( Image(fig1a),Move(40,40), ) ) );

fig1b = NewImage("J:\MAJUG\Feb 2009\AJR192_Fig1b.gif");
w = New Window(
"Figure 1B (Reprinted with permission from the American Journal of Roentgenology
)",
   gb = Graph Box(
      Frame Size( 250, 250 ), X Scale( 0, 80 ), Y Scale( 0, 80 ),
      << Add Image( Image(fig1b), Move(40,40),) ) );
```

Figures 4. Representative CT Images of Basion-dens measurements of Normal (4a) and Abnormal (4b) measurements

4a (< 12 mm, non-CCDI, Controls) 4b (≥ 12 mm, CCDI, Study Grp)



Reprinted with permission from the *American Journal of Roentgenology*

The **Add Image ()** function brings in images (in .gif,.bmp, .jpg, or .png format) into a Graph Window. Options to specify bounds, position, and resizing are also available. For more information, visit John Ponte's blog [3].

JMP's Logistic Regression and Recursive Partition platforms analyzed the data and found the measures which differentiated the CCDI and non-CCDI groups.

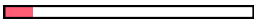
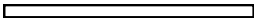
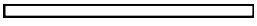
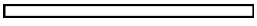
Logistic regression in Table 5 showed the multivariate likelihood ratio effect test results produced by the JSL script in Appendix 2. Four of the seven variables were statistically significant. The ROC Area Under Curve (0.9338) depicted the number of true positives (sensitivity of finding CCDI patients) plotted against the false positives (finding patients without CCDI).

Table 5. Logistic Regression Multivariate Effect Likelihood Ratio Tests (ROC Area Under Curve = 0.9338, Rsquare (U) = 0.589)

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Midline occiput-C1 spinolaminar distance	1	1	2.16376147	0.1413
Basion-dens tip distance	1	1	4.57674764	0.0324*
Condylar sum	1	1	4.61037524	0.0318*
Midline C1-C2 spinolaminar line	1	1	4.47573276	0.0344*
ADI	1	1	7.31248656	0.0068*
Vertical C1-dens line	1	1	1.55111237	0.2130
Basion-posterior axial line	1	1	1.09613058	0.2951

Table 6 listed the significant variables from the recursive partitioning. Recursive partitioning splits the data into groups of input variables (Column Contributions) that best separated or partitioned CCDI and non-CCDI patients. The column contributions showed how each variable contributed to the overall model fit based on the number of splits and the size of the G^2 (likelihood chi-square) statistic that considers all the possible ways of splitting the data to maximize the difference between patients with and without CCDIs. The ROC Area Under Curve was 0.9701.

Table 6. Column Contributions ROC Area Under Curve = 0.9701, Rsquare = 0.781

Term	Number of Splits	G^2	G^2
Midline occiput-C1 spinolaminar distance	1	7.376762	
Basion-dens tip distance	2	62.226078	
Condylar sum	0	0.000000	
Midline C1-C2 spinolaminar line	0	0.000000	
ADI	1	21.502085	
Vertical C1-dens line	0	0.000000	
Basion-posterior axial line	1	4.129581	

Cross-validation was used to avoid overfitting the data. Overfitting means the model fits noise as well as signal to the extent the model has poor predictive accuracy and low generality with future data. Cross-validation randomly divides the data into two parts, training and test. The training set builds a fitting model to its data. The test set assesses the predictive performance of the fitting model on the other data. A validation column divides the data into training and test datasets. The validation column was formed by using JMP's **Random Uniform ()** function to randomly assign values of 0 or 1 for the Training or Test (or Validation) datasets. For more information, visit Valerie Hyde's blog [4].

The JSL code below adds two columns to a Data Table with formulas. The Randno column uses the **Random Uniform ()** function to create a column of random numbers between 0 and 1. The Validation column assigns records of the data to training (0) if the Randno value is below two-thirds. Otherwise, Validation is assigned the test value of one.

```
New Column( "Randno",Numeric, Continuous, Format( "Best", 12 ),
          Formula( Random Uniform() ) );
New Column( "Validation",Numeric, Continuous,Format( "Best", 12 ),
              Formula( If( :Randno < 2 / 3, 0, 1 ) ) );
```

JMP provides the K-fold method of cross-validation. JMP Pro (new in version 9) offers the Bootstrap Forest and Boosted Trees methods of cross-validation.

K-fold validation randomly divides the original data into K subsets of roughly equal size (JMP's default is K=5). The training data are used to fit a predictive model on the K-1 subsets. The held-out K-th subset is used to test (or validate) the predictive accuracy of the fitting model. Each of the K subsets is used to validate the model fit of the other subsets of data. The model with the best validation statistics (e.g., ROC Area under Curve, Rsquare (U), etc.) is reported along with the fit of the original data.

The choice of validation method depends upon the study objectives, analytical, computing resources, amount of data available, and other considerations.

K-fold cross-validation is useful when there are limited data available for model fitting (e.g., small sample sizes) and all the data are needed to perform the validation process. Boosted Trees and Random Forests are useful when there are large amounts of data available and it is important to compare several types of fitted models. The separate, fitted models receive fair assessments based on the measures of fit with the test or validation datasets.

Bootstrap Forests build many trees and averages the predicted values to produce a final set of predictions. Each tree grows from different random samples (with replacement) of the observations. .

Table 7 showed the results of applying the Boosted Forests method to the data (100 trees). The Misclassification rates of the fitted model of the Training and Test (Validation) dataset were 0.1071 (10.7%) and 0.0333 (3.33%), respectively.

Table 7. Bootstrap Forest Training and Validation Summary

Bootstrap Forest for Group

Specifications

Target Column:	Group	Training rows:	56
Validation Column:	Validation	Validation rows:	30
		Test Rows	0
Number of trees in the forest:	100	Number of terms:	7
Number of terms sampled per split:	4	Bootstrap samples:	56
		Minimum Splits Per Tree:	10
		Minimum Size Split:	5

Overall Statistics

Measure	Training	Validation	Definition
Entropy RSquare	0.5699	0.4991	1-Loglike(model)/Loglike(0)
Generalized R-Square	0.7281	0.6473	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.2978	0.3060	∑ -Log(ρ[j])/n
RMSE	0.2838	0.2775	√∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.2380	0.2543	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1071	0.0333	∑ (ρ[j]≠ρMax)/n
N	56	30	n

Confusion Matrix

Actual / Training	Controls	Study Grp
Controls	29	0
Study Grp	6	21

Predicted

Actual / Validation	Controls	Study Grp
Controls	21	0
Study Grp	1	8

Predicted

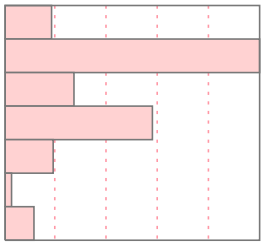
The significant columns contributing classifying patients with CCDI and without CCDI are listed in Table 8.

Table 8. Significant Columns from the Bootstrap Forest Partition

Bootstrap Forest for Group

Column Contributions

Term	Number of Splits	G^2
Midline occiput-C1 spinolaminar distance	55	244.2199
Basion-dens tip distance	97	1,336.4640
Condylar sum	65	365.3579
Midline C1-C2 spinolaminar line	59	777.4717
ADI	22	260.5483
Vertical C1-dens line	15	35.0827
Basion-posterior axial line	38	158.9983



Boosted Trees is a second method in JMP Pro for identifying significant variables. Boosted Tress builds a large, final tree by fitting random samples (with replacement) from sequences of smaller trees. Each of the smaller trees use assigned weights to cases from previous trees (called reweighting). Higher weights are assigned to misclassified cases . Smaller weights are assigned to correctly classified cases for each iteration of earlier built trees. The validation column assesses how many stages (or layers) to fit, not to exceed a specified number of stages. The misclassification rates for the Boosted Forest method in Table 9 were 0 (0%) and 0.0333 (3.33%) for the training and test datasets, respectively.

Table 9. Boosted Tree Training and Validation Summary

Boosted Tree for Group

Specifications

Target Column:	Group	Number of training rows:	56
Validation Column:	Validation	Number of validation rows:	30
Number of Layers:	17		
Splits Per Tree:	4		
Learning Rate:	0.1		
Overfit Penalty:	0.0001		

Overall Statistics

Measure	Training	Validation	Definition
Entropy RSquare	0.9632	0.8427	1-Loglike(model)/Loglike(0)
Generalized R-Square	0.9825	0.9114	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0255	0.0961	∑ -Log(ρ[i])/n
RMSE	0.0376	0.1684	√ ∑(y[i]-ρ[i])²/n
Mean Abs Dev	0.0247	0.0660	∑ |y[i]-ρ[i]|/n
Misclassification Rate	0.0000	0.0333	∑ (ρ[i]≠ρMax)/n
N	56	30	n

Confusion Matrix

Actual		Predicted	Actual		Predicted
Training	**Controls**	**Study Grp**	**Validation**	**Controls**	**Study Grp**
Controls	29	0	Controls	20	1
Study Grp	0	27	Study Grp	0	9

Table 10 listed the significant variables contributing to the G^2 statistics using the Boosted Tree method.

Table 10. Significant Columns from the Boosted Tree

Boosted Tree for Group

Column Contributions

Term	Number of Splits	G^2
Midline occiput-C1 spinolaminar distance	6	152.07994
Basion-dens tip distance	9	260.93506
Condylar sum	16	400.19005
Midline C1-C2 spinolaminar line	7	150.24413
ADI	18	356.11561
Vertical C1-dens line	11	177.87996
Basion-posterior axial line	1	41.98016

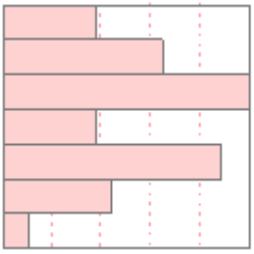
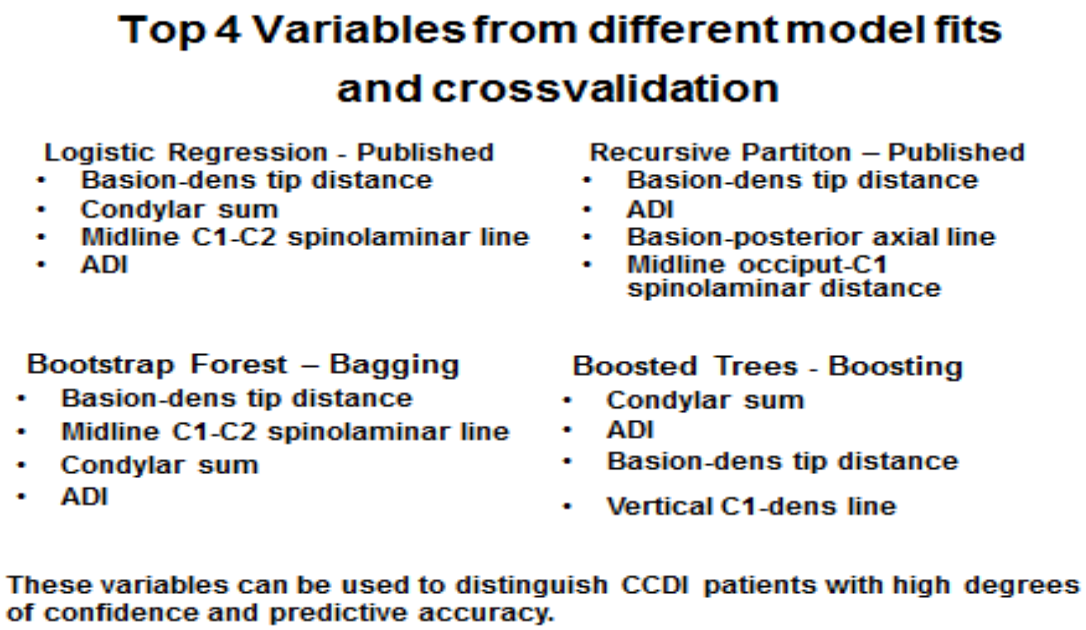


Table 11 summarized the top four variables from the four model fits and cross-validations.

Table 11. Top 4 Variables from the Model Fitting and Validation Methods that Identified Patients with and Without CCDIs



CONCLUSION

From a Google Scholar search of citations of the published investigations above on July 26, 2011, Case Study 1 was cited 12 times and Case Study 2 was cited four times. The number of studies where JMP has been used increased from two in 2007-2008 to 11 since 2009 (3 pending publication; 5 undergoing DSMB and IRB review or approval; 3 in investigational design and sample-size determination stages). Statistical methods in Neurosurgery and Radiology are growing with increased momentum. They have been standard in pharmaceutical clinical trials for years. JMP continues to play a key role that makes it possible to assess the impact of innovative surgical and imaging techniques of interest to neurosurgeons and radiologists.

APPENDICES

Appendix 1. Scripts for generating the data table, contingency table analyses, odds ratios and plot for Case Study 1

```
// Scripts used for Case Study 1 SDG following DC


//Step 1: Script to produce Contingency Table of Fisher's
// Exact, Pearson, and Likelihood Tests with
//Odds Ratios

New Table( "Neurosurg Focus 26 2009 E8 Table 1",
       Add Rows( 48 ), New Property( "Contingency ",
       New Window( "Neurosurg Focus 26 2009 E8 Table 1- Contingency of SDG by Var",
                    V List Box( Contingency( Y( :SDG ), X( :Var ), Freq( :Count ),
                                 Contingency Table( Count( 1 ), Total %( 0 ),
                                         Col %( 1 ), Row %( 0 ), Expected( 0 ),

Deviation( 0 ), Cell Chi Square( 0 ) ), Mosaic Plot( 0 ), Odds Ratio( 1 ),
Where( :G == 1 ) ), Contingency( Y( :SDG ), X( :Var ), Freq( :Count ),
Contingency Table( Count( 1 ), Total %( 0 ), Col %( 1 ), Row %( 0 ),
Expected( 0 ), Deviation( 0 ), Cell Chi Square( 0 ) ), Mosaic Plot( 0 ),
Odds Ratio( 1 ), Where( :G == 2 ) ), Contingency( Y( :SDG ), X( :Var ),
Freq( :Count ), Contingency Table( Count( 1 ), Total %( 0 ), Col %( 1 ),
```

```
Row %( 0 ), Expected( 0 ), Deviation( 0 ), Cell Chi Square( 0 ) ),
Mosaic Plot( 0 ), Odds Ratio(1), Where( :G == 3 ) ),
Contingency( Y( :SDG ), X( :Var ), Freq( :Count ), Contingency Table(
Count( 1 ), Total %( 0 ), Col %( 1 ), Row %( 0 ), Expected( 0 ),
Deviation( 0 ), Cell Chi Square( 0 ) ), Mosaic Plot( 0 ), Odds Ratio( 1 ),
Where( :G == 4 ) ), Contingency( Y( :SDG ), X( :Var ), Freq( :Count ),
Contingency Table( Count( 1 ), Total %( 0 ), Col %( 1 ), Row %( 0 ),
Expected( 0 ), Deviation( 0 ), Cell Chi Square( 0 ) ), Mosaic Plot( 0 ),
Odds Ratio( 1 ), Where( :G == 5 ) ),
Contingency( Y( :SDG ), X( :Var ), Freq( :Count ),
Contingency Table( Count( 1 ), Total %( 0 ), Col %( 1 ), Row %( 0 ),
Expected( 0 ), Deviation( 0 ), Cell Chi Square( 0 ) ), Mosaic Plot( 0 ),
Odds Ratio( 1 ), Where( :G == 6 ) ),
Contingency( Y( :SDG ), X( :Var ), Freq( :Count ),
Contingency Table( Count( 1 ), Total %( 0 ), Col %( 1 ), Row %( 0 ),
Expected( 0 ), Deviation( 0 ), Cell Chi Square( 0 ) ), Mosaic Plot( 0 ),
Odds Ratio( 1 ), Where( :G == 7 ) ),
Contingency( Y( :SDG ), X( :Var ), Freq( :Count ),
Contingency Table( Count( 1 ), Total %( 0 ), Col %( 1 ), Row %( 0 ),
Expected( 0 ), Deviation( 0 ), Cell Chi Square( 0 ) ), Mosaic Plot( 0 ),
Odds Ratio( 1 ), Where( :G == 8 ) ),
Contingency( Y( :SDG ), X( :Var ), Freq( :Count ),
Contingency Table( Count( 1 ), Total %( 0 ), Col %( 1 ), Row %( 0 ),
Expected( 0 ), Deviation( 0 ), Cell Chi Square( 0 ) ), Mosaic Plot( 0 ),
Odds Ratio( 1 ), Where( :G == 9 ) ),
Contingency( Y( :SDG ), X( :Var ), Freq( :Count ),
Contingency Table( Count( 1 ), Total %( 0 ), Col %( 1 ), Row %( 0 ),
Expected( 0 ), Deviation( 0 ), Cell Chi Square( 0 ) ),
Mosaic Plot( 0 ), Odds Ratio( 1 ), Where( :G == 10 ) ),
Contingency( Y( :SDG ), X( :Var ), Freq( :Count ),
Contingency Table( Count( 1 ), Total %( 0 ), Col %( 0 ), Row %( 1 ),
Expected( 0 ), Deviation( 0 ), Cell Chi Square( 0 ) ),
Mosaic Plot( 0 ), Odds Ratio( 1 ), Where( :G == 11 ) ),
Contingency( Y( :SDG ), X( :Var ), Freq( :Count ),
Contingency Table( Count( 1 ), Total %( 0 ), Col %( 0 ), Row %( 1 ),
Expected( 0 ), Deviation( 0 ), Cell Chi Square( 0 ) ), Mosaic Plot( 0 ),
Odds Ratio( 1 ), Where( :G == 12 ) ) ) ) ),
New Column( "G", Numeric, Nominal, Format( "Best", 15 ),
             Set Property( "Value Ordering", {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}
),
             Set Values(
                    [1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6,
6, 6,
                    7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10, 10, 11, 11, 11,
11, 12,
                    12, 12, 12]
             ),
             Value Labels(
                    {1 = "MVA", 2 = "Fall", 3 = "5mm Shift", 4 = "Cpr/Obl Basal
Cistern", 5
                     = "IVH", 6 = "Sub Hemorrhage", 7 = "Diffuse injury", 8 =
                    "Evacuated Mass", 9 = "Early Surgery", 10 = "Delayed Surgery", 11
=
                    "GOS 1-3", 12 = "GOS 4-5"}
             ),
             Use Value Labels( 1 )
       ),
       New Column( "Var",
             Character,
             Nominal,
             Set Values(
                    {"MVA", "MVA", "No MVA", "No MVA", "Fall", "Fall", "No Fall", "No
Fall",
```

```
                    " >5mm Shift", " >5mm Shift", "<5mm Shift", "<5mm Shift",
                    "Cpr/Obl Basal Cistern", "Cpr/Obl Basal Cistern",
                    "No Cpr/Obl Basal Cistern  ", "No Cpr/Obl Basal Cistern ", "IVH",
"IVH",
                    "No IVH", "No IVH", " Sub Hemorrhage", " Sub Hemorrhage",
                    "No Sub Hemorrhage", "No Sub Hemorrhage", "Diffuse injury",
                    "Diffuse injury", "No Diffuse injury", "No Diffuse injury",
                    "Evacuated Mass", "Evacuated Mass", "No Evacuated Mass",
                    "No Evacuated Mass", "Early Surgery", "Early Surgery",
                    "No Early Surgery", "No Early Surgery", "Delayed Surgery",
                    "Delayed Surgery", "No Delayed Surgery", "No Delayed Surgery",
"GOS 1-3","GOS 1-3", "No GOS 1-3", "No GOS 1-3", "GOS 4-5", "GOS 4-5","No GOS 4-5",
"No GOS 4-5"} ) ),
      New Column( "SDG",
            Character,
            Nominal,
            Set Values(
                  {"With ", "Without", "With ", "Without", "With ", "Without", "With
",
                  "Without", "With ", "Without", "With ", "Without", "With ",
"Without",
                  "With ", "Without", "With ", "Without", "With ", "Without", "With
",
                  "Without", "With ", "Without", "With ", "Without", "With ",
"Without",
                  "With ", "Without", "With ", "Without", "With ", "Without", "With
",
                  "Without", "With ", "Without", "With ", "Without", "With ",
"Without",
                  "With ", "Without", "With ", "Without", "With ", "Without"}
            )
      ),
      New Column( "Count",
            Numeric,
            Continuous,
            Format( "Best", 10 ),
            Set Values(
                  [31, 14, 8, 15, 3, 10, 36, 19, 14, 15, 25, 14, 33, 20, 6, 9, 13,
8, 26,
                  21, 24, 21, 15, 8, 30, 15, 9, 14, 9, 24, 30, 5, 16, 15, 23, 14,
23, 14,
                  16, 15, 15, 15, 24, 14, 24, 14, 15, 15]
            )
      ),
      New Column( "Total",
            Numeric,
            Nominal,
            Format( "Best", 10 ),
            Set Values(
                  [68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68,
68, 68,
                  68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68,
68, 68,
                  68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68, 68]
            )
      ),
            New Column( "Proportion",
            Numeric,
            Nominal,
            Format( "Best", 10 ),
            Formula( :Count / :Total ),
            Lock( 1 )
```

```
	)
);


//Step 2: Use Summarized to output the SDG risk factors

Data Table( "Neurosurg Focus 26 2009 E8 Table 1" ) <<
Summary( Group( :G ), output table name( "G " ) );

// Output data table from right-clicking the Odds Ratios
// from the Contingency of SDG by Var and Make Combined
// Data Table

New Table( "Odds Ratios",
	Add Rows( 12 ),

	New Column( "Odds Ratio",
		Numeric,
		Continuous,
		Format( "Best", 12 ),
		Set Values(
			[4.15178571428571, 0.158333333333333, 0.522666666666667, 2.475,
1.3125,
			0.60952380952381, 3.11111111111111, 0.0625, 0.649275362318841,
			1.54017857142857, 0.583333333333333, 1.71428571428571]
		)
	),
	New Column( "Lower 95%",
		Numeric,
		Continuous,
		Format( "Best", 12 ),
		Set Values(
			[1.43099261071723, 0.0388567477133398, 0.196349391314248,
			0.765980027563526, 0.458451844040857, 0.215680954084746,
			1.09765010045446, 0.0184910582917767, 0.24654889209022,
			0.584850962233707, 0.220478864364843, 0.647937887113009]
		)
	),
	New Column( "Upper 95%",
		Numeric,
		Continuous,
		Format( "Best", 12 ),
		Set Values(
			[12.0457118284541, 0.645176087031029, 1.39129763843898,
7.99710799181637,
			3.75755114172139, 1.72254094457705, 8.81793965278334,
0.211250753654115,
			1.70983731681098, 4.05599064559603, 1.54335781236016,
4.5355821424462]
		)
	)
);


// Join the G and Contingency Odds ratios
// to form Odds Ratios and 95% CIs Data Table
Data Table( "Odds Ratios" ) << Join(
	With( Data Table( "G " ) ),
	SelectWith( :G ),
	Select( :Odds Ratio, :Lower 95%, :Upper 95% ),
	By Row Number,
	Preserve main table order( 1 ),
	Output Table( "Odds Ratios and 95% CIs" ) ) ;
```

```
dtor = Data Table ( "Odds Ratios and 95% CIs" );

dtor <<
Overlay Plot( X( :G ), Y( :Odds Ratio, :Lower 95%, :Upper 95% ), Needle( 1 ),
       :Odds Ratio( Overlay Marker( 8 ) ),
       :Lower 95%( Overlay Marker( 25 ) ),
       :Upper 95%( Overlay Marker( 24 ) ),
       SendToReport(
              Dispatch(
                     {},
                     "Overlay Plot",
                     OutlineBox,
                     {Set Title(
                            "Overlay Plot of Odds Ratios and 95% Confidence Intervals
of Risk Factors" )}),
              Dispatch(
                     {},
                     "102",
                     ScaleBox,
                     {Format( "Best", 10 ), Add Ref Line( 1, Solid, "Black" )}
              ),
              Dispatch(
                     {},
                     "101",
                     ScaleBox,
                     {Min( 0 ), Max( 13 ), Inc( 1 ), Minor Ticks( 0 ),
                     Rotated Labels( "Angled" )}
              ),
              Dispatch(
                     {},
                     "Overlay Plot Graph",
                     FrameBox,
                     {Frame Size( 927, 202 ), Marker Size( 4 ), Line Width Scale( 2 )}
              ) ) );
```

Appendix 2. Scripts for generating the data table, logistic regression and recursive partitioning for Case Study 2

```
dtcam = Open("J:\MAJUG\Feb 2009\CAM_AJR192_Data.jmp");

//Logistic Regression on original data
dtcam << Fit Model(
       Y( :Group ),
       Effects(
              :Name( "Midline occiput-C1 spinolaminar distance" ),
              :Name( "Basion-dens tip distance" ),
              :Condylar sum,
              :Name( "Midline C1-C2 spinolaminar line" ),
              :ADI,
              :Name( "Vertical C1-dens line" ),
              :Name( "Basion-posterior axial line" )
       ),
       Personality( Nominal Logistic ),
       Run(
              Positive Level( "Study Grp" ),
              Likelihood Ratio Tests( 1 ),
              Wald Tests( 0 ),
              Odds Ratios( 1 ),
              ROC Curve( 1 ),
              Confidence Intervals( 1 ),
              Profiler(
                     1,
                     Term Value(
```

```
                        Name( "Midline occiput-C1 spinolaminar distance" )(4.2),
                        Name( "Basion-dens tip distance" )(9.5),
                        Condylar sum( 6.18 ),
                        Name( "Midline C1-C2 spinolaminar line" )(8.89),
                        ADI( 2.811 ),
                        Name( "Vertical C1-dens line" )(0.572),
                        Name( "Basion-posterior axial line" )(5.446) ) ) ),
      SendToReport(
            Dispatch(
                  {},
                  "Parameter Estimates",
                  OutlineBox,
                  {Set Title( "Parameter Estimates" )}
            ),
            Dispatch(
                  {},
                  "Effect Likelihood Ratio Tests",
                  OutlineBox,
                  {Set Title(
                        "Logistic Regression Effect Likelihood Ratio Tests (ROC
Area Under Curve = 0.9338, Rsquare (U) = 0.589) "
                  )}
            ),
            Dispatch( {"Odds Ratios"}, "Range Odds Ratios", OutlineBox, {Close( 1 )}
),
            Dispatch(
                  {},
                  "Receiver Operating Characteristic",
                  OutlineBox,
                  {Set Title( "Receiver Operating Characteristic" )} ) ) );
// Partition on original data
dtcam << Partition( Y( :Group ), X(
            :Name( "Midline occiput-C1 spinolaminar distance" ),
            :Name( "Basion-dens tip distance" ),
            :Condylar sum, :Name( "Midline C1-C2 spinolaminar line" ),
            :ADI, :Name( "Vertical C1-dens line" ),
            :Name( "Basion-posterior axial line" ) ),
      Small Tree View( 1 ), Leaf Report( 1 ), Column Contributions( 1 ),
      ROC Curve( 1 ), Criterion( "Maximize Significance" ),
      Initial Splits( :Name( "Basion-dens tip distance" ) >= 9.4,
            {:Name( "Basion-posterior axial line" ) >= 3.5},
            {:ADI < 3, {}, {:Name( "Midline occiput-C1 spinolaminar distance" ) < 7,
            {:Name( "Basion-dens tip distance" ) < 7.1}}} ),
      SendToReport(
            Dispatch( {}, "Partition", FrameBox( 2 ), {Frame Size( 400, 87 )} ),
            Dispatch( {}, "Column Contributions",
                  OutlineBox,
                  {Set Title(
                        "Column Contributions ROC Area Under Curve = 0.9701,
Rsquare = 0.781" )} ) ) );
// Partition for Bootstrap Forests
dtcam << Partition( Y( :Group ),
      X( :Name( "Midline occiput-C1 spinolaminar distance" ),
            :Name( "Basion-dens tip distance" ), :Condylar sum,
            :Name( "Midline C1-C2 spinolaminar line" ),:ADI,
            :Name( "Vertical C1-dens line" ),
            :Name( "Basion-posterior axial line" ) ),
      Validation( :Validation ), Method( Bootstrap Forest ),
      Portion Bootstrap( 1 ), Number Terms( 4 ),
      Number Trees( 100 ), Go,
      SendToReport(
            Dispatch( {}, "Cumulative Validation", OutlineBox, {Close( 0 )} ),
            Dispatch( {}, "Per-Tree Summaries", OutlineBox, {Close( 0 )} ) ) ) ;
```

```
// Partition for Boosted Trees
dtcam << Partition( Y( :Group ),
       X(     :Name( "Midline occiput-C1 spinolaminar distance" ),
              :Name( "Basion-dens tip distance" ), :Condylar sum,
              :Name( "Midline C1-C2 spinolaminar line" ), :ADI,
              :Name( "Vertical C1-dens line" ),
              :Name( "Basion-posterior axial line" ) ),
       Validation( :Validation ), Method( Boosted Tree ),
       Splits Per Tree( 4 ), Number of Layers( 17 ),
       Learning Rate( 0.1 ), Early Stopping( 1 ), Go,
       SendToReport(
              Dispatch( {}, "Cumulative Validation", OutlineBox, {Close( 0 )}),
              Dispatch( {}, "Tree Views", OutlineBox, {Close( 0 )} ) ) );
```

REFERENCES

1. Aarabi B., Chesler D., Maulucci C., Blacklock T., Alexander M. (2009) “Dynamics of subdural hygroma following decompressive craniectomy: a comparative study,” *Neurosurgical Focus*, 26, E8.
2. Chang, W., Alexander M.T., Mirvus S.E. (2008),: “Diagnostic Determinants of Craniocervical Distraction Injury in Adults,” *American Journal of Roentgenology*, 192, 52-58.
3. John Ponte (May 23, 2011), “How to Add an Image to a Graph in JMP 9”, http://blogs.sas.com/jmp/index.php?/archives/429-How-to-Add-an-Image-to-a-Graph-in-JMP-9.html.
4. Valerie Hyde (July 6, 2010), “Train, Validate and Test for Data Mining in JMP“, http://blogs.sas.com/jmp/index.php?/archives/324-Train,-Validate-and-Test-for-Data-Mining-in-JMP.html.

ACKNOWLEDGMENTS

I thank Bizhan Aarabi, David Chesler, Christopher Maulucci, Tiffany Blacklock, Wilbur Chang, and Stu Mirvis for their contributions to this paper. I thank Lucia Ward-Alexander for her review and editorial assistance.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Melvin Alexander
Phone: (410) 458-7129
E-mail: Melvin.alexander@verizon.net