2012

Rescoring the Montana Risk Assessment Instrument: A Comparison of Methods

Patrick David McKay
The University of Montana

Rescoring the Montana Risk Assessment Instrument:

A Comparison of Methods

By

Patrick David McKay

B.A., The University of Montana, Missoula, MT, 2012

Thesis

presented in partial fulfillment of the requirements

for the degree of

Master of Arts

in Sociology

The University of Montana

Missoula, MT

Spring 2012

Approved by:

Dr. J.B. Alexander Ross

Graduate School

Dr. Dusten Hollist, Chairperson

Sociology Department

Dr. James Burfeind

Sociology Department

Dr. Douglas Dalenberg

Economics Department

McKay, Patrick, M.A., Spring 2012   *Sociology*

Rescoring the Montana Risk Assessment Instrument: A comparison of methods

Chairperson or Co-Chairperson: Dusten Hollist, Ph.D.

The purpose of this research is to examine alternate ways to add meaningful weights to the risk factors on the Montana Risk Assessment Instrument (RAI). An evaluation is made which compares the predictive accuracy of a revised scoring system compared to the one that is currently in use. The data for this analysis is taken from 299 Montana juveniles who were administered the RAI after an offense, between January 01, 2009 to December 31, 2010. The results are based on a Burgess model, a linear probability model, and a logistic regression model. The findings suggest that all three models increased the predictive accuracy of the RAI. The Burgess model and the logistic model showed the greatest improvement. When considering both predictive accuracy and practical usability, the Burgess model for rescoring the RAI was found to be the best approach.

The small sample size was a limitation in this research which may have affected the statistical significance of the risk factors found on the RAI when using linear probability and logistic regression. Inconsistencies found between counties when collecting data was another limitation in this research. Finally, the inability to find a continuous outcome variable forced this research to use a linear probability model instead of a linear regression model. Future research to increase the predictive accuracy of the RAI must concentrate on three major topics. First, it must be a priority to find appropriate risk factors for the RAI. Second, continue research that will determine the best approach to add meaningful weight to risk factors. Finally, examine the cut point on the RAI to eliminate the most false positive and false negative predictions.

Acknowledgment

I would like to thank the entire Sociology department for their guidance and support during my four years as an undergraduate students and two years as a graduate student. I have had the privilege of getting to know many individuals in the department and could not imagine a greater group of people.

I owe a special thanks to the members of my committee, Dusten Hollist, Jim Burfeind, and Doug Dalenberg, for their continued advice and support throughout my research. I can honestly say, without the hours of guidance from the chair of my committee, Dusten Hollist, I would not be where I am now.

I would also like to thank my friends and family who have always been there for me. I am truly blessed to have such great people in my life.

Contents

INTRODUCTION

Risk assessment instruments are used in the juvenile justice system in an attempt to eliminate the subjective nature of many decisions in the processing of youth. This research is concerned with the Montana Pre-adjudicatory Detention Risk Assessment Instrument (RAI). The RAI is comprised of a list of weighted risk factors that when summed provide an objective means to determine whether youth should be released or be placed into detention before their initial court appearance.

As part of the Juvenile Detention Alternative Initiative (JDAI), four counties in Montana have employed the use of the RAI: Missoula, Hill, Yellowstone, and Cascade. The RAI was developed to consider two primary factors: 1) the likelihood that a released youth will appear for a subsequent judicial preceding and 2) the likelihood that a released youth will commit a new offense during the period of risk between release from detention and adjudication.

The primary research objective is to use statistical methods, instead of human judgment, to revise the weights that are assigned to items on the RAI. An evaluation will be made which compares the predictive accuracy of a revised scoring system compared to the one that is currently in use. The research hypothesis is twofold: 1) Rescoring items on the Montana RAI using the Burgess Method, linear probability model, and logistic regression, will increase the predictability of a youth receiving a new citation within a year of release from detention. 2) When considering both predictive accuracy and practical usability, the Burgess Method will yield the best revised RAI for Montana juveniles.

The information that follows is organized into four sections. Section one will provide background on the JDAI and the research that is currently being done on the Montana RAI. It

also reviews prior studies that have focused on building and scoring risk assessment tools. Descriptive statistics are presented in section two. These cover key elements of the sample and how it is divided into a construction and a validation sample. In this section, a thorough discussion of the RAI and the methods used to create the revised scores are discussed. The results of the rescoring procedures are presents in section three. The results are organized into two parts: first, a comparison of the construction sample to the validation sample, second, a comparison of the validation sample to the original RAI sample. In the fourth and final section, discussion of the results, limitations in the current investigation, and recommendations for future research are addressed.

SECTION 1: JDAI and Literature on Risk Assessment Instruments

<u>Juvenile Detention Alternative Initiative (JDAI)</u>

Since its origins in 1992, the Juvenile Detention Alternatives Initiative (JDAI) has been a key part of the Anne E. Casey Foundation's mission toward detention reforms across the United States. According to data presented on the Casey Foundation Webpage (www.aecf.org), at the time that this report was written there were 100 JDAI sites in 24 states and the District of Columbia. The initiative was designed to support the vision that all youth involved in the juvenile justice system have opportunities to develop into healthy, productive adults.

In Montana there are four pilot counties (Cascade, Hill, Missoula, and Yellowstone) that were initially involved in the movement toward alternatives to secure confinement of juveniles. In each of the JDAI counties, a coordinator is selected as the point of contact. The JDAI coordinator then works with local Juvenile Justice System stakeholders to identify resources and develop strategies to promote the use of alternatives to secure confinement and detention reform.

Risk assessment instruments play an important role in detention reform. These instruments are a key piece in the process of evaluating juveniles who have been arrested for a detainable offense to determine the need for confinement in secure detention. The instruments are expected to be based on objective criteria (e.g. criminal background) and uniformly applied to all juveniles who have committed a detention eligible offense.

Research for the Anne E. Casey Foundation suggest using the "consensus design" over the more formal "prediction method" to create risk assessment instruments (Steinhart 2006). The consensus design relies on local stakeholders in the juvenile justice system to use their professional knowledge to select and add weight to risk factors. The consensus design is posited to be tailored to local policy, laws, and the youth population. Alternatively, the formal

prediction method uses statistical tests on data collected from juveniles to find risk factors and

associated weight. The formal prediction methodology is considered to be time consuming,

expensive and inapplicable once created (Steinhart 2006). Once the instrument has been

implemented, it can then be formally validated on a sample of released youth to determine the

relative rate of success and failure on a particular outcome (e.g., new citation).

<p align="center">Montana RAI Validation Study</p>

Over the past year, researchers at The University of Montana have been conducting a

validation study on the RAI. Their analysis focuses on two dimensions associated with the RAI.

The first of these pertains to racial and cultural sensitivity in assessing offender risk. The second

pertains to public safety outcomes associated with the behavior of youth who are released from

detention; specifically, whether a new offense occurred resulting in a misdemeanor or felony

citation and whether the youth failed to appear for an initial court appearance after release. To

achieve these objectives, the following three research questions were examined:

1. Is the RAI being administered impartially and in a manner that it assesses juvenile
   offender "risk" in a culturally and racially sensitive manner?
2. Did the juveniles reoffend while on release status during the period of risk?
3. Did the juvenile fail to appear for the initial court appearance following release from
   detention?

Findings from the RAI validation study are as follows: Agreement between the RAI

indicated decision and actual decision was the most common outcome found which encompassed

52% (333 of 621) of the decisions. Cases involving minority juveniles were more likely to result

in agreement than those involving White juveniles. Overrides down, where the actual decision

<p align="center">4</p>

was less punitive than the RAI indicated decision, occurred in 27.7% (172 of 621) of the total outcomes. Overrides up, where the actual decision was more punitive than the RAI indicated decision, occurred in 15% (93 out of 621) of the outcome decisions, most of which (78 out of 93; 83.8%) involved cases pertaining to White juveniles. Override decision from a detention alternative to detention were most likely to occur in cases involving White juveniles (63 of 93 cases; 67.7%).

Overall, the RAI had a failure rate of 12.2%, which is just over the 10% threshold recommended by the JDAI. The RAI results indicated good performance for felony citations with a failure rate of 1.5% (2 out of 130 cases). The RAI results for misdemeanors citations were just over the recommended failure rate at 10.8% (14 out of 130 cases). Interestingly, the overall failure rate for juveniles that received a detention override was notably higher at 22.0% (28 out of 127 cases) (Hollist, Coolidge, Delano, Greenwood, King, McLean, Mckay, Burfeind, Harris, and Doyle; Forthcoming).

Risk Assessment Instruments

Risk assessment instruments are not limited to the Juvenile Justice System. There are a variety of instruments that measure for several types of risk, used in different professional fields. For example, there are risk assessment instruments developed specifically for nonsexual violence (Violence Risk Appraisal Guide; Harris Rice, and Quinsey 1993), for sexual violence (Rapid Risk Assessment for Sex Offense Recidivism; Hanson 1997), and for general recidivism (Level of Service Inventory-Revised; Andrews and Bonta 1995). Other instruments in common use include the psycho-diagnostic tool most commonly used to assess psychopathy (Psychopathy Checklist-Revised; Hare 1991), and the Self-Appraisal Questionnaire designed to predict violent

and nonviolent offender recidivism (Loza and Green 2003). While these are just a few examples, they provide some insight to just how large the field of risk assessment research is.

<div align="center">Creating Risk Assessment Instruments Using the Formal Prediction Method</div>

Researchers have been studying formal prediction methodologies for over 80 years. In 1928 E. W. Burgess created one of the first risk assessment instruments using what would later be called, the Burgess Method (Burgess 1928). This is a linear additive model that looks at several risk predicting variables. For each risk variable that applies to an individual one point is added to their total score. Thus, the more points an individual scores on the instrument the more likely the individual is to act out the risk behavior being predicted (e.g., recidivate). Since the creation of the Burgess Method, researchers have been examining ways to increase the predictability of risk behavior by finding both alternate models that predict risk, and ways to add meaningful weight to risk predicting variables.

Literature based on the comparison of statistical methods use a variety of tests in an attempt to develop the most predictive risk assessment instrument. These include: the Burgess Method (Gottfredson and Gottfredson 1980; Silver, Smith, and Banks 2000; Gottfredson and Snyder 2005; Caulkins, Cohen, Gorr, and Wei 1996; Kirby 1954), multiple linear regression (Simon 1972; Gottfredson and Snyder 2005; Gottfredson and Gottfredson 1980; Aguinis and Gottfredson 2010), logistic regression (Silver, Smith, and Banks 2000; Gottfredson and Snyder 2005; Thomas, Leese, Walsh, McCrone, Moran, Burns, Creed, Tyrer, and Fahy 2004), classification tree method (Thomas et al. 2004), iterative classification (Silver, Smith, and Banks 2000), recursive partitioning (Silver, Smith, and Banks 2000), neural network model (Caulkins, Cohen, Gorr, and Wei 1996), multiple models tool (Silver and Chow-Martin 2002), configural model (Gottfredson and Gottfredson 1980), multivariate contingency (Gottfredson and

<div align="center">6</div>

Gottfredson 1980), discriminant analysis (Gottfredson and Snyder 2005), clustering methods (Gottfredson and Snyder 2005) and bootstrap methods(Gottfredson and Snyder 2005). Of these, the Burgess Method, multiple linear regression, and logistic regression are the most common methods found in the literature.

A common finding is that there are few differences and most models perform equally well when predicting risk (Simon 1972, Gottfredson and Gottfredson 1980; Caulkins et al. 1996; Gottfredson and Snyder 2005). The most advanced techniques for weighting variables still do not significantly outperform simple tests such as the Burgess method (Gottfredson and Snyder 2005; Simon 1972, Silver, Smith, and Banks 2000). Gottfredson and Snyder (2005) discuss the importance of simplicity, face validity, and flexibility as key advantages for researchers to consider when a risk assessment instrument is applied in the field for use. If a tool measured objectively outperforms all other traditional tools it is useless if it does not have the ability to be applied in the field which requires more subjective judgment.

<u>Terminology</u>

There is consistent terminology in all risk assessment research and it might help to clarify some of these more common terms. *Outcome* refers to the anticipated event (e.g., receipt of a new misdemeanor or felony citation). Failure to appear was not included in the outcome variable because of the small amount of juveniles who were indicated to have failed to appear. *Period of Risk* is the span of time after release from detention that the juvenile is eligible to receive a misdemeanor or felony citation. The period of risk for this research is one year after release from detention. The term *failure* is used to indicate a juvenile received a new citation in the period of risk. The term *success* is used to indicate a juvenile did not receive a new citation in the period of

risk. The term *baserate* refers to the percentage (or proportion) of cases that are indicated as a failure. *Risk factors* are referring to the measurable case characteristics found on the RAI that are being used in an attempt to predict the outcome. While *RAI* is a broad term used in all risk assessment research, the RAI that is analyzed in this report is the Montana Pre-adjudicatory Detention Risk Assessment Instrument. The term *thresholds* and *cut points* are used to indicate the number of points a juvenile must score on the RAI to move from the release option to the detention alternative option or the secure detention option.

SECTION 2: Data and Methods

Data

The sample is comprised of a total of 299 Montana juveniles who were administered the

RAI after an offense between January 01, 2009 to December 31, 2010. Juveniles that were

released from detention within five days of placement are included in the sample. This amount

of time was allowed because five days is the maximum amount of time a juvenile can remain in

detention and still have an indicated "release" outcome. This also allowed for an increased

sample size. Scores collected from the original RAIs that were administered to youth in Hill,

Yellowstone, Missoula, and Cascade counties are the basis for the analysis that will follow. All

other data was collected from the Juvenile Court Activity Tracking System (JCATS). The

JCATS is used in all counties in Montana to keep juvenile detention records. These records

include information such as identification number, case notes from the probation officers, social

information, family information, and information about current and past offenses.

The demographics of the total sample are presented in Table 1 below. The sample is

69.7% (210) White, 21.1% (63) American Indian or Alaska Native, .3% (1) Asian, 3.3% (10)

African American, and 5.6% (15) Hispanic or Latino. 66.6% (199) of the juveniles are male and

the remaining 33.4% (100) are female. The juvenile's ages range from 10 to 18 with the

majority being 14 to 16.

Table 1: Sample Descriptive Statistics

Race	Male Frequency	Male %	Female Frequency	Female %	Success %	Failure %
White	146	69.5%	64	30.5%	47.6%	52.4%
American Indian or Alaska Native	34	54.0%	29	46.0%	28.6%	71.4%
Asian	1	100.0%	0	0.0%	100.0%	0.0%
Black or African American	7	70.0%	3	30.0%	30.0%	70.0%
Hispanic or Latino	11	73.3%	4	26.7%	53.3%	46.7%
Total	199	66.6%	100	33.4%	43.5%	56.5%

N= 299

The data in Table 1 shows that 56.5% (169) of the juveniles in the total sample received a new citation for a misdemeanor or felony offense in the period of risk (Failure). Of those who reoffended in the period of risk 79.8% (135) received misdemeanor citations while the remaining 20.2% (34) were felony citations. White juveniles are almost split in half for those that succeeded and those that failed. Alternatively, American Indians or Alaska Native failed at a much higher percentage with 71.4% (45) failing and only 28.6% (18) succeeding. An accurate comparison cannot be made with the sample of Asian, Black or African American, and Hispanic or Latino juveniles due to low numbers in each group in the total sample.

To help determine the validity of the new weights on the RAI a construction and a validation sample must be created. The construction sample is used to build the new RAIs. The validation sample is used after the RAIs have been built to examine the variation in prediction accuracy from the original construction sample to the validation sample.

Construction and validation samples were generated by randomly dividing the sample of juveniles in half. Dividing the sample in half is the recommended procedure for small sample size RAI research by Gottfredson and Snyder (2005). To do this, all cases were sorted by the date the RAI was administered and then every other case was selected to create approximately equal samples. Silver, Smith and Banks (2000) used a similar procedure explaining this

provided control over the effects of time on rates and correlates of receipts of new offenses. Those juvenile who were included in the construction or validation more than once were removed from the sample except for the case with the earliest RAI administration date. These juveniles were removed from the sample so each individual case would be separate and independent from all other cases in the sample. However, an analysis was completed with these juveniles in the samples and no significant differences were found. Once complete, the construction sample consists of a total of 151 juveniles and the validation sample has a total of 148 juveniles.

The construction and validation sample are very similar to the full sample. In the construction sample males make up 63.4% (97) and females make up 36.6% (56). There are a few more males in the validation sample (69.5%, 104) and a few less females (30.5%, 46). The construction and validation sample were almost identical when comparing the race of the juveniles. There are 69.9% (107) White juveniles in the construction sample and 69.5% (105) White juvenile in the validation sample. There are 20.9% (32) Native Americans or Alaskan Natives in the construction sample compared to 21.2% (32) in the validation sample. 58% (87) of the juveniles in the construction sample received a new citation in the period of risk and 55.4% (82) of the juveniles in the validation sample received a new citation in the period of risk.

<u>The Montana Pre-adjudicatory Detention Risk Assessment Instrument</u>

A copy of the RAI is included in Appendix A. It was modified from Virginia's Detention Assessment Instrument (DAI). In accordance with JDAI recommendations, Virginia's DAI was developed using the consensus design by the Department of Juvenile Justice (DJJ) with assistance from the National Council of Crime and Delinquency (NCCD) and a group of key

stakeholders from across the state of Virginia (Steinhart 2006). Reiner, Miller and Gangal

(2007) conducted a validation on Virginia's DAI and found that it passed the requirements set by

the JDAI.

The RAI is almost identical to the Virginia instrument. Virginia's DAI has two categories

that are not on the RAI: History of Failure to Appear, and History of Escape/Runaway. Also, the

RAI has two categories that are not on Virginia's DAI: Current Warrant or Pickup Order, and

Warrant History. In creating the RAI, the scores were increased for "Most Serious Offense

Alleged in Current Referrals" category, "Additional Offense Alleged in Current Referrals"

category, and "Supervision Status" category when compared to the Virginia instrument. By

increasing the scores in these categories, there are 62 total possible points on the RAI versus 43

points on the DAI. However, the number of points needed to exceed the initial release indicated

option to a more restrictive indicated decision is the same on both instruments.

The RAI consists of seven predictor categories which generate a total score. The total

score is then classified into one of three categories as the "indicated decision."

- 0-9 Release

- 10-14 Detention Alternative

- 15+ Secure Detention

The RAI also allows for an "Override Justification" where the Juvenile Probation Officer can

detain a youth when the RAI recommends release (aggravating override), or can release when

the RAI recommends detained (mitigating override), based on individual case by case discretion.

The seven risk factor categories on the RAI are based on a juvenile's current offense and

prior offense history. To analyze individual risk factor weights, each of the seven variables were

broken down into 19 dummy variables. Table 2 shows the descriptive statistics for each risk

factor based on the device construction sample. All risk factors are coded as dichotomies.

Pearson Correlation was used to analyze each risk factor's correlation with the outcome (new

citation in period of risk). Pearson Correlation ranges from -1 to 1. Scores close to 1 or -1

indicate a strong correlation and scores closest to 0 indicate a weak correlation. A risk factor

with a negative Pearson Correlation indicates that juveniles who have that risk factor are less

likely to receive a new citation in the period of risk than those juvenile who do not have that risk

factor. These risk factors were selected to be on the RAI based on prior knowledge of their

association to the outcome. As is apparent in the Pearson Correlation, only three risk factors are

significantly correlated with the outcome measure. One of these "Most serious offense alleged

in current referral, felony against persons" is negatively correlated with the outcome. It is

expected that all risk factors would be positively correlated to the outcome. However, seven of

the 19 risk factors were found to have a negative Pearson Correlation scores.

Table 2: Descriptive Statistics and Correlations With Outcome Measures for Construction Sample (n=151)

Risk Factors	M	SD	Pearson Correlation
Warrant or Pickup Order	0.16	0.37	.153
Most serious offense alleged in current referral felonies against persons	0.05	0.22	-.216[**]
Most serious offense alleged in current referral other felony	0.08	0.28	-.119
Most Serious offense alleged in current referral Misd. against person	0.34	0.48	.017
Most serious offense alleged in current referral other Misd.	0.24	0.43	-.055
One or more additional current felony offenses	0.01	0.11	.099
One or more additional misd. or violation of prob or parole offenses	0.36	0.48	-.019
Prior admissions of guilt to two or more felony offenses	0.02	0.14	-.070
Prior admissions of guilt to one felony offense	0.10	0.30	-.003
Prior admissions of guilt for two or more misd. or status offenses	0.44	0.50	.280[**]
Prior admissions of guilt for two or more probation or parole violations	0.01	0.08	.070
Prior admission of guilt for any misd or status	0.16	0.36	-.067
One or more pending referrals for a felony offense	0.06	0.24	.159
Two or pending referrals for other offenses	0.11	0.32	.136
One pending referral for other offense/offenses	0.07	0.26	.137
intensive or close supervision	0.10	0.31	.240[**]
Formal release conditions /on probation/ on parole	0.38	0.49	.099
Warrant history: Two or more warrants	0.04	0.19	.106
Warrant history: One Warrant	0.04	0.19	.106

NOTE: ** coefficients are significant at p < .01

Montana RAI Scores and Statistics

Figure 1 shows the distribution of juveniles who succeeded and failed (received a new misdemeanor or felony citation) by the total score received on the RAI. There are a total of 299 juveniles included in Figure 1. Of these for who the RAI indicated decision was *release* (scores 0-9) 60% succeeded and 40% received a new citation during the period of risk. For juveniles whose RAI indicated score suggested *detention alternative* (scores 10-14) 43.0% succeeded 56.96% failed due to behavior that resulted in a new misdemeanor or felony citation. 34.3% of all juveniles who were released from detention within five days of confinement despite a RAI indicated decision to *detain* (score 15 points or higher) were successful and did not receive a new citation during the period of risk. In contrast 65.7% failed as a result of a new felony or misdemeanor citation.

Figure 1:
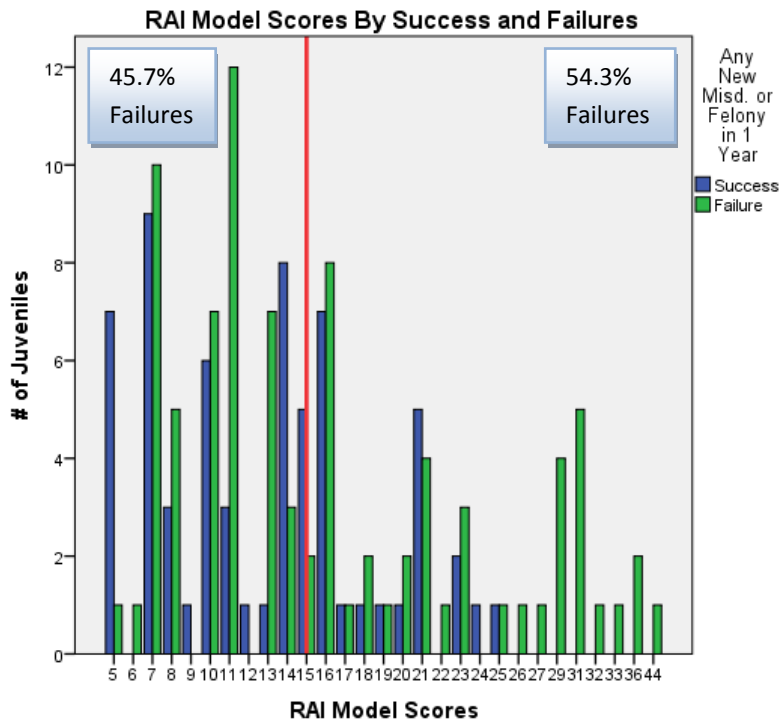
In figure 1 the red line is placed on the 15 point mark indicating the secure detention threshold. Of the 169 juveniles that failed 54.3% were above the 15 point secure detention threshold and 45.7% were below the threshold. According to RAI cut points, the juveniles to the right of the red line should have been detained and those to the left should have been released or received an alternative to detention. If the risk factors were weighted correctly and were valid predictors for the outcome (receipt of a new citation in the period of risk) it would be expected that most of the failures (green bars) would be to the right of the red line, and most of the successes (blue bars) would be to the left of the red line. This pattern does exist to some extent, however, it is clear that the observed pattern diverges from what was expected where failures are lowest at the lower end of the RAI continuum, higher in the middle, and highest after the 15 point threshold.

<div align="center">Statistical Procedures for Device Rescoring</div>

Device #1: Based on the Burgess Method.

In the Burgess Method rescoring the risk factors were coded as dichotomies where the value of 1 indicates the presence of the characteristic known to be associated with failure, and 0 indicates the youth did not have that characteristic associated with failure. Total Burgess scores were computed for each juvenile by summing across the 19 items that comprise the 7 risk domains on the RAI. For example, a juvenile that has two of the 19 risk factors would receive a total RAI score of 2 points.

Table 3: Distribution of Success and Failures Based on Burgess Scoring

Burgess Score	Success		Failure		Total
	n	%	n	%	
1	22	68.75%	10	31.25%	32
2	8	34.78%	15	65.22%	23
3	23	39.66%	35	60.34%	58
4	10	40.00%	15	60.00%	25
5	1	9.09%	10	90.91%	11
6	0	0.00%	1	100.00%	1
7	0	0.00%	1	100.00%	1
Total:	64	42.38%	87	57.62%	151

Table 3 shows the distribution of juveniles that succeeded and failed based on their total Burges scores. It is expected that those who score low on this test should have a higher likelihood of success and those who score high should have a higher likelihood of failure. To some extent this pattern is apparent in the distribution. 68.75% (22) of those who scored a 1 on the test succeeded while 31.25% (10) failed. Additionally, 29% (11) of those juveniles who scored at least a 4 succeeded compared to 71% of the juvenile who failed.

Device #2: Based on the Linear Probability Model.

This procedure provides a linear equation similar to the Burgess method for calculating a total risk score. However, instead of arbitrarily assigning a value of 1 for each risk variable, the linear probability model provides an estimated weight for each variable. The weight for each variable comes from the unstandardized regression coefficient. Since all risk factors are coded as dichotomies the unstandardized regression coefficient is appropriate to use. The coefficient measures the change in probability of a juvenile failing when the risk factor is present while holding all other variables constant. To provide a simple form for scoring purposes the regression coefficient is rounded to two decimal places and multiplied by 100. This technique was used by Gottfredson and Snyder 2005. Multiplying the coefficients by 100 is necessary for

each factor to have its own unique weight. Similar to the Burgess Method, total scores are computed by summing across all 19 risk factors that comprise the RAI.

Table 4 contains the linear probability results for the risk factors in the construction sample. Four variables are included in the linear probability model so their effects are held constant providing a more accurate coefficient estimate. The four variables being held constant are: county the youth was living in, the juvenile's race, if the juvenile was detained or released, and the juvenile's sex. The "County" variable is used in the model to hold constant the differences between the four counties in Montana. The "Race" and "Sex" variables are used in the model to hold constant the differences between males and females and their race on the RAI. Finally, the variable "Was Youth Detained" is used in the model to hold constant the differences between those juveniles who were officially detained and then released in the five day period from those juveniles who were never officially detained in that five day period.

Table 4: Linear Probability Results (construction sample)

Risk Factors:	Unstandardized Coefficient	Standard Error	Sig.	Weighted Risk Score
Warrant or pickup order	.024	.196	.904	-
Most serious offense alleged in current referral felony against persons	-.374	.197	.060	-
Most serious offense alleged in current referral other felony	-.093	.168	.580	-
Most serious offense alleged in current referral misd. against persons	.136	.129	.295	14
Most serious offense alleged in current referral other misd.	-.036	.121	.766	-
One or more additional misd. or violation of prob. or parole offenses	-.145	.098	.142	-
Prior admissions of guilt to two or more felony offenses	-.024	.291	.934	-
Prior admissions of guilt to one felony offense	.170	.153	.269	17
Prior admissions of guilt for two or more misd. or status offenses	.314	.109	.005	31
Prior admission of guilt for any misd. or status offense	.222	.121	.068	22
One or more pending referrals for a felony offense	.260	.160	.106	26
Two or more pending referrals for other offenses	.165	.126	.193	17
One pending referral for other offense	.134	.150	.374	13
Intensive or close supervision	.362	.175	.041	36
Formal release conditions /on probation/ on parole	.086	.117	.460	9
Warrant history: Two or more warrants	.166	.219	.449	17
Warrant history: One warrant	-.214	.221	.333	-
Missoula Dummy	.283	.133	.035	-
Hill dummy	-.120	.182	.510	-
Yellowstone dummy	.073	.100	.465	-
Race: White	-.247	.088	.006	-
Was youth detained	-.236	.088	.008	-
Sex	.082	.080	.309	-

R= .550

R²= .303

Two risk factors found on the RAI are not present in this results table because they showed perfect collinearity with the outcome variable which will give a biased standard error, low significance, and an inflated coefficient. Perfect collinearity means that the risk factor is perfectly associated to the outcome variable. The two omitted risk factors are "One or more additional felony offenses," and "Prior admissions of guilt for two or more probation or parole violations." The risk factor "Warrant or pickup order" will not be included in the newly weighted RAI based on the extremely low significance associated with that variable. Surprisingly, six of the risk factors showed a negative correlation to the outcome measure when all variables were held constant. These risk factors will not be included in the newly weighted RAI since it would not be logical to include negative weights with these risk factors.

The new reweighted RAI will consist of 10 of the original 19 risk factors. This model is called the full linear probability model. It was originally thought that all, or most of the variables would be statistically significant. To test the differences a separate analysis was conducted using the four most statistically significant risk factors (Prior admission of guilt for two or more misdemeanor or status offenses, Prior admission of guilt for any misdemeanor or status offense, Intensive or close supervision, and One or more pending referrals for a felony offense). This model is called the significant linear probability model.

Figures 3 and 4 (below) show the distribution of successes and failures along the total RAI score for the significant linear probability model and the total linear probability mode. Again, it is expected that there will be a higher frequency of failures (green) on the right hand side of these tables, and a higher frequency of successes (blue) on the left hand side. This pattern is visible in both of these models but the higher variation in the full linear probability model may make this the preferred model. While some of the error has improved from the original RAI distribution, it is apparent there are risk factors missing from the RAI that would further separate those who succeed from those who fail.

Figures 3 and 4:
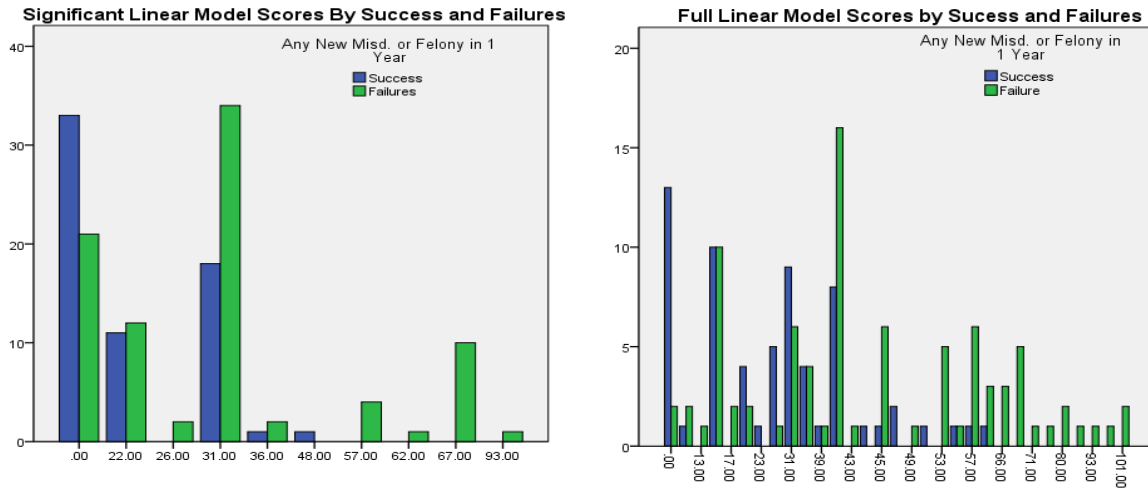
Device #3: Based on Logistic Regression.

Similar to the linear probability model, logistic regression allows for each of the risk factors to have a unique weight. The weight for the logistic model is based on the mean marginal effects for each of the risk factors. Again, the marginal effects were rounded to two decimal places and multiplied by 100 to give each risk factor a simple, unique weighted score. When turned into a percentage, the marginal effects provide a good approximation to the amount of percentage point change in the outcome variable that will be produced by a 1 unit change in the risk factors holding all else constant. Total scores were calculated by summing all risk factors.

Table 5: Logistic Regression Results (Construction sample)

	UC	Marginal	S.E	Sig.	OR	Weighted Risk Score
Warrant or pickup order	-0.38	-0.06	1.22	0.75	0.68	-
Most serious offense alleged in current referral felony against persons	-3.24	-0.52	1.93	0.09	0.04	-
Most serious offense alleged in current referral other felony	-0.56	-0.09	1.16	0.63	0.57	-
Most serious offense alleged in current referral misd. against persons	0.56	0.09	1.09	0.60	1.75	9
Most serious offense alleged in current referral other misd.	-0.40	-0.06	1.00	0.69	0.67	-
One or more additional misd. or violation of prob or parole offenses	-0.94	-0.15	0.59	0.11	0.39	-
Prior admissions of guilt to two or more felony offenses	0.09	0.01	1.50	0.95	1.09	-
Prior admissions of guilt to one felony offense	0.87	0.14	0.84	0.30	2.38	14
Prior admissions of guilt for two or more misd. or status offenses	1.82	0.29	0.63	0.00	6.20	29
Prior admission of guilt for any misd. or status offense	1.26	0.20	0.69	0.07	3.51	20
One or more pending referrals for a felony offense	1.90	0.30	1.25	0.13	6.66	30
Two or more pending referrals for other offenses	1.03	0.17	0.77	0.18	2.79	17
One pending referral for other offense	0.81	0.13	0.93	0.39	2.24	13
Intensive or close supervision	2.74	0.44	1.37	0.05	15.41	44
Formal release conditions /on probation/ on parole	0.41	0.07	0.67	0.55	1.50	7
Warrant history: Two or more warrants	1.05	0.17	1.40	0.45	2.86	17
Warrant history: One warrant	-1.54	-0.25	1.63	0.34	0.21	-
Missoula Dummy	1.15	0.18	0.89	0.19	3.17	-
Hill dummy	-1.10	-0.18	0.56	0.48	0.33	-
Cascade dummy	-1.33	-0.06	0.95	0.25	0.67	-
Race: White	-1.52	-0.24	0.54	0.01	0.22	-
Was youth detained	-1.33	-0.21	0.53	0.01	0.26	-
Sex	0.46	0.08	0.46	0.31	1.59	-

NOTE: UC= Unstandardized Coefficients: OR= Odds Ratios.

Table 5 presents the logistic regression results for the risk factors in the construction sample. Similar to the linear probability model there are four variables being held constant (County, Race, Sex, and if the youth was detained or not). Also, the same two variables were omitted due to perfect collinearity (one or more additional felony offenses and prior admissions of guilt for two or more probation or parole violations). Five items will not be included in the logistic weighting because they were found to be negatively correlated with the outcome. The factor "Prior admission of guilt to two or more felony offenses" will also not be included as the result of the low observed significance level.

The ten remaining risk factors will be reweighted with logistic regression. Like the linear probability model discussed above, a separate RAI will be made of the four most significant risk factors (Prior admission of guilt for two or more misdemeanor or status offenses, Prior admission of guilt for any misdemeanor or status offense, Intensive or close supervision,

and One or more pending referrals for a felony offense). This model will be called the significant logistic model in the results that follow.

Figures 5 and 6 show the distribution of failures and successes by the logistic weighted RAI score for the significant logistic regression model and the full logistic regression model. The logistic distributions look similar to the linear probability model distributions. Once again, there is middle zone with a high percentage of both successes and failures.

Figures 5 and 6:
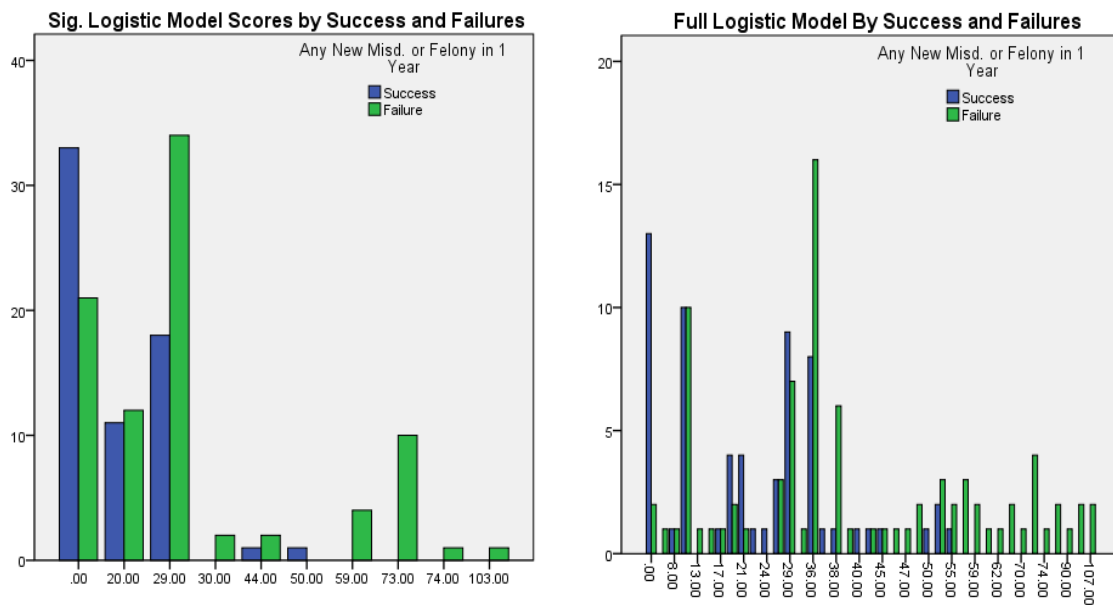
<div align="center">Defining thresholds</div>

A youth can fall into one of three categories on the RAI: release, detention alternative, or secure detention. To determine where these thresholds fall, a formula from Silver, Smith, and Banks (2000) is used. The researchers created a formula for high risk threshold and one for a low risk threshold based on the sample baserate. Silver et al. identified these thresholds by doubling and halving the odds of recidivism using the following formulas:

High= $\dfrac{2*\frac{p}{1-p}}{1+(2*\frac{p}{(1-p)})}$

Low= $\dfrac{.5*\frac{p}{1-p}}{1+(.5*\frac{p}{(1-p)})}$

> For the high-risk and low-risk threshold, respectively: where high represents the sample baserate of recidivism associate with a doubling of odds of recidivism for the total sample; low represents the baserate of recidivism associated with a halving of the odds of recidivism for the total sample; p is the sample baserate and p/(1-p) is the odds associated with the sample baserate (Silver, Smith and Banks 2000;746-747)

Silver et al. used this formula to distinguish between two groups (low and high) then reanalyzed the middle sample that was left. Instead of using the sample as a reanalysis sample it will be used as the detention alternative category. The baserate for the construction sample is .58 meaning that 58% of the juvenile's in the construction sample received a new citation for a misdemeanor or felony in the period of risk.

Once the baserate has been placed in each of these formulas a proportion for low risk and a proportion for high risk is calculated. The low risk calculated proportion is .41 and the high risk calculated proportion is .734. To determine where these proportions fall, each of the models were put into a crosstab table with RAI scores in the rows and Success and Failure in the columns. By using the cumulative marginal total percent in the crosstab table a line can be drawn at the bottom 41% of the juveniles then another line at the top 73% of juveniles. The RAI scores that correspond to the low and high proportions are the thresholds.

Table 6: Calculated Thresholds

	Release	Det. Alt	Detain
Burgess	0-2	3	4+
Full Log	0-27	28-35	36+
Sig. Log	0-19	20-28	29+
Full Lin	0-31	32-44	45+
Sig. Lin	0-22	23-30	31+

Table 6 presents the cut points for each model. In all models the bottom 41% of the distribution of juveniles fall in the release option, 32% of the middle distribution fall in the detention alternative option, and the top 27% of the distribution fall in the detain option. After creating the thresholds for each model, initial comparison can then be made between the construction sample and the original RAI.

Table: 7 Threshold Results (Construction vs. Original RAI)

Model	Success and Low Score		Failure and High Score		Success and Middle Score		Failure and Middle Score		Correct		Incorrect		λ
Original RAI:	16.05%	48	30.77%	92	11.37%	34	15.05%	45	61.87%	185	38.13%	114	12.30%
Burgess:	19.87%	30	17.88%	27	15.23%	23	23.18%	35	60.93%	92	39.07%	59	7.80%
Full Log Model:	25.17%	38	37.09%	56	5.96%	9	5.3%	8	67.55%	102	32.45%	49	25%
Sig. Log Model:	21.85%	33	35.76%	54	7.28%	11	7.95%	12	65.56%	99	34.44%	52	18.75%
Full Lin. Model:	28.48%	43	25.83%	39	9.27%	14	14.57%	22	68.87%	104	31.13%	47	26.56%
Sig. Lin. Model:	29.20%	44	34.40%	52	0%	0	1.30%	2	64.90%	98	35.10%	53	17.20%

Table 7 shows the results for the construction sample compared to the original RAI on the full sample. Red indicates the lowest percent found in the column and blue represents the highest percent. The significant linear model shows the greatest percentage of juveniles (29.2%) classified as low risk that did not receive a new citation in the period of risk. The full logistic model classified the highest percent of juveniles (37.1%) as high risk that did also receive a new citation in the period of risk. Overall the full linear model was the most accurate in its predictions classifying 68.87% of the youth into correct risk groups. Alternatively, the Burgess Model performed the worst, classifying 60.93% of the youth into correct risk groups. According

to the calculated lambda, every model has a higher reduction in prediction error compared to the

original RAI except for the Burgess model. The Burgess model has the greatest percent of

juveniles classified in the middle group for both success and failures and it seems that this is

largely responsible for the reduction in classification accuracy for the Burgess model.

SECTION 3: Results

A unique aspect of this approach is the ability to compare the RAI indicated result to an actual RAI. Most literature on the development of risk assessment instruments do not have this ability and must stop the analysis after comparing validation models to the construction model. Model validity can be tested for by comparing the construction model to the validation model then the validation model can be compared to the actual RAI to determine if predictive accuracy has been increased. The results are presented below.

Construction vs. Validation Sample

Table 8 compares the results for the validation sample when compared to the construction sample. The following observations may be drawn. Three of the validation samples unexpectedly outperformed the construction samples. The significant linear model in the validation sample outperformed the construction sample by 1.99 percentage points. The full logistic model in the validation sample outperformed the construction sample by .69 percentage points. Finally, the Burgess model in the validation sample outperformed the construction sample by 9.34 percentage points. While it is a positive sign that the validation samples are just as accurate as the construction sample, the large variation in the Burgess model is unexpected and may warrant concern. All other models were within 2 percentage points of each other showing evidence of their validity from one sample to the next.

Table 8: Construction vs. Validation

Model	Success and Low Score		Failure and High Score		Success and Middle Score		Failure and Middle Score		Correct		Difference (Pct. Points)
Con: Burgess	19.87%	30	17.88%	27	15.23%	23	23.18%	35	60.93%	92	9.34
Val: Burgess	28.38%	42	20.27%	30	11.49%	17	21.62%	32	70.27%	104	
Con: Full Log Model	25.17%	38	37.09%	56	5.96%	9	5.3%	8	67.55%	102	0.69
Val: Full Log Model	25.00%	37	33.11%	49	8.11%	12	10.14%	15	68.24%	101	
Con: Sig. Log Model	21.85%	33	35.76%	54	7.28%	11	7.95%	12	65.56%	99	-0.7
Val: Sig. Log Model	23.65%	35	37.17%	55	6.08%	9	4.05%	6	64.86%	96	
Con: Full Lin. Model	28.48%	43	25.83%	39	9.27%	14	14.57%	22	68.87%	104	-1.98
Val: Full Lin. Model	28.38%	42	25.00%	37	8.78%	13	13.51%	20	66.89%	99	
Con: Sig. Lin. Model	29.20%	44	34.40%	52	0%	0	1.30%	2	64.90%	98	1.99
Val: Sig. Lin. Model	29.97%	44	35.81%	53	0%	0	1.35%	2	66.89%	99	

Second, the significant linear model in the validation sample correctly predicted the highest percent of juveniles in the low risk category. The significant logistic model in the validation sample correctly predicted the highest percent of juveniles in the high risk category. Finally, the Burgess Model in the validation sample was the most accurate of those examined, model correctly placing 70.3% of the juveniles.

Overall, the construction and validation samples have similar placement percentages. This is a positive result that indicates empirical validity. Another test of empirical validity can be calculated by determining the Pearson Correlation for each model's totals to the outcome (receipt of a new citation in the period of risk) and comparing their correlations. This method was borrowed from Gottfredson and Snyder (2005). However, instead of using the Point Biserial Correlation the Pearson Correlation was used. This is demonstrated in table 9.

Table 9: Correlation of Prediction Scores with Outcomes

Prediction Method	Pearson Correlation		Shrinkage
	Construction	Validation	
Original RAI	0.293		-
Burgess	0.278	0.429	-15.1%
Full Log	0.358	0.38	-2.2%
Sig. Log	0.42	0.29	12.9%
Full Lin	0.412	0.392	2.0%
Sig. Lin	0.359	0.31	4.9%
NOTE: All Models are Significant at p < .01			

According to the Pearson Correlation all models including the original RAI are statistically significant at the 99% level. In the construction sample, the significant logistic model has the strongest correlation to the outcome with a Pearson Correlation of .42. Overall, the Burgess method in the validation sample has the strongest correlation out of all models with a Pearson Correlation of .429. To determine shrinkage, the Pearson Correlation from the construction sample is subtracted from the Pearson Correlation from the validation sample for all models.

According to Gottfredson and Snyder (2005:27), "A smaller amount of shrinkage might give greater confidence that the validity of the prediction method, and the classification procedure derived from it, will hold up on repeated applications." Models with shrinkage closest to zero are the most empirically valid. As table 9 shows, the Burgess model and the significant logistic model have the greatest amount of shrinkage indicating that these models may be the least empirically valid. Alternatively, the full logistic model, the full linear model and the significant linear model all have relatively little shrinkage indicating these models may be more consistent when repeated on other juvenile samples.

Validation vs. Original RAI

Table 10 compares the validation sample results to the original RAI results. The following observations may be drawn from the findings. Every model outperforms the original RAI according to both the correct predictions percentages and the calculated lambda. According to the correct prediction column, the Burgess model has the greatest increase in accuracy with an increase of 8.4 percentage points from the original RAI. For the rest of the models predictive accuracy increased by at least 5 percentage points with the exception of the significant logistic model which only outperformed the original RAI by 2.99 percentage points.

Table 10: Validation vs. Original RAI

Model	Correct	Incorrect	Comparison predicted correct to Original RAI (Pct. Points)	λ
Original RAI:	61.87% 185	38.13% 114	-	12.30%
Burgess:	70.27% 104	29.73% 44	8.4	33.30%
Full Log Model:	68.24% 101	31.76% 47	6.37	28.79%
Sig. Log Model:	64.86% 96	35.14% 52	2.99	25.76%
Full Lin. Model:	66.89% 99	33.11% 49	5.02	25.76%
Sig. Lin. Model:	66.89% 99	33.11% 49	5.02	25.76%

The lambda reduction in errors exhibits a greater degree of separation between the created models and the original RAI. All models reduced the error of predicting the probability of receiving a citation by twice the percentage of the original RAI. The Burgess model has the greatest calculated Lambda and indicates by using the Burgess model the amount of error predicting if a juvenile would receive a new misdemeanor or felony citation is reduced by 33.3%, while the original RAI reduced the error by12.3%.

Figures 7 and 8 demonstrate the difference between the original RAI and the Burgess model on the validation sample. It is apparent that the Burgess model is more successful at

separating those who have received a new citation from those who did not. These results provide

the initial evidence to indicate that rescoring the RAI could be beneficial for making a more

predictive tool than the current version of the RAI.
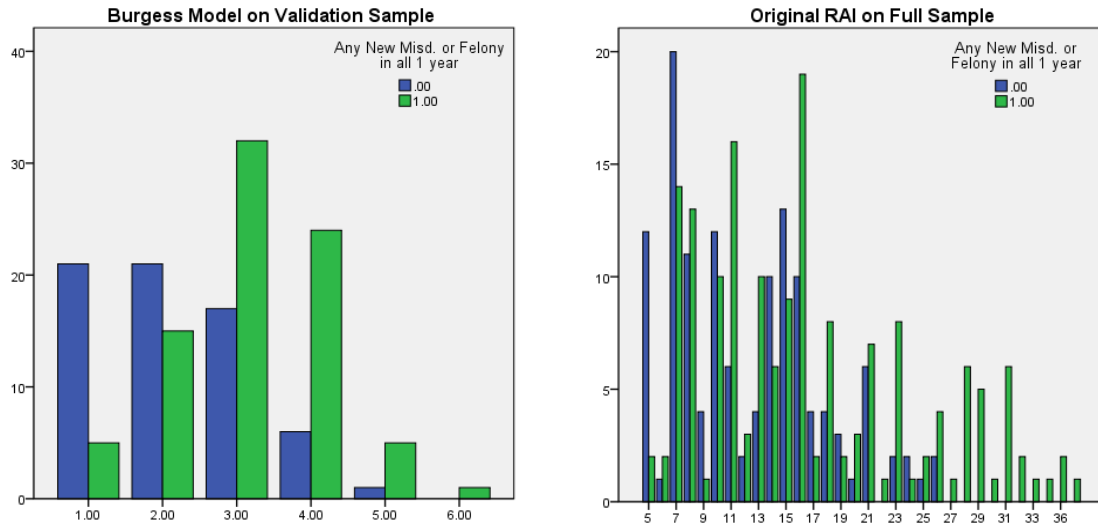
Figures 7 and 8:

Table 11 presents a comparison of the models. To calculate the overall rank, each model

was placed in order from best to worst performing on three comparison categories (Compared to

construction, Empirically Valid, and Compared to Original RAI). The rank for each category was

summed and the lowest score was found to be the model with the best performance overall.

Table 11: Model Overall Ranking

Rank	Compared to Construction	Empirically Valid	Compared to Original RAI	Overall
1	Burgess Model	Full Linear Model	Burgess Model	Burgess Model
2	Sig. Linear Model	Full Log Model	Full Log Model	Full Log Model
3	Full Log Model	Sig. Linear Model	Full Linear Model	Full Linear Model
4	Sig. Log Model	Sig. Log Model	Sig. Linear Model	Sig. Linear Model
5	Full Linear Model	Burgess Model	Sig. Log Model	Sig. Log Model
NOTE: The Burgess Model and the Full Log Model are tied for first and second				
NOTE: The Sig. Linear Model and the Full Linear Model are tied for third and Fourth				

Ranking the models using these categories is not an accurate depiction of their abilities and is only meant for a visual comparison. When the validation models are compared to the construction models the Burgess model is on top. It completely outperformed the construction sample which was unexpected. The full linear model and the full logistic model are the most empirically valid, based on the shrinkage found when comparing the Pearson Correlation. Once again, the Burgess model is the least empirically valid model. In the comparison of the validation to the original RAI the Burgess model was once again on top, followed by the full logistic model. Based on these results the Burgess model and the full Logistic model were tied for best overall performance. The Burgess model is questionable based on its performance from construction to validation sample. Alternatively, the full logistic model is found to be empirically valid and scored well in both the comparison to the construction sample and the original RAI.

SECTION 4: Conclusion and Discussion

Conclusion

The goal of this research was to find a way to add meaningful weights to the scoring system that is currently used to assess risk of the probability of receiving a new citation during the period of risk. Both hypotheses examined in the investigation were partially supported. Rescoring items on the RAI using the Burgess Method, linear probability, and logistic regression improved the ability to correctly predict the likelihood of a new misdemeanor or felony citation within a year of release from detention. The Burgess model in the validation sample showed the most prediction accuracy followed by the logistic model and then the linear model. The examination based only on items that were statistically significant outperformed the original RAI, however, these are based on only a few risk factors. Other items need to be added to the models before their performance can be accurately analyzed.

When considering both predictive accuracy and practical usability, there is evidence to suggest a weighting system based on the Burgess method is the best option. When considering usability, the Burgess model will always be the most simplistic model. While the Burgess method was the most accurate on the validation sample, it is unknown how this model will perform on another sample in the future based on the tests for empirical validity. However, based on the findings and the simplicity of the model, it is the recommendation of this research that the Burgess Method is the preferred model to use to weight items on the RAI when compared to a logistic model, a linear model and the model currently being used.

The sample size in this research is a limitation with many implications. Risk factors that were not statistically significant may very well have been if a larger sample had been used. Also,

some of the risk factors may be negatively correlated to the outcome because of the way the sample was selected. Risk factors that measured for serious behavior such as, "Most serious offense alleged in current referral was a felony against persons," may be largely absent from the present investigation as they would have been detained longer than the five day period that was employed due to the seriousness of the offense.

Inconsistencies found between counties in the way practitioners document juvenile information was another limitation. In the JCATS, it was found that counties differed in the way they document cases making data collection more difficult in some counties. It became apparent when collecting RAI scores that a total RAI score could be different depending on the county and probation officer that was conducting the test. Moreover, RAIs are intended to be filled out and used as a tool to help probation officers and the courtroom workgroup come to a decision on what to do with a juvenile. It was found in some counties, RAIs were filled out after the fact making their usefulness obsolete and possibly skewing the data. As discussed earlier, in an attempt to correct these inconsistencies the county variables were held constant for the linear probability and logistic regression models.

To determine the impact the variable "was youth detained" had on the linear probability model and logistic regression model, a separate regression was run for both with this variable omitted. No significant differences were found for either the coefficients or the statistical significance level for either model. While this variable did not have a large impact on the results it was found to be statistically significant in the logistic model and very close to statistically significant in the linear probability model. This indicates that there may be differences between those youth that were detained and those that weren't that may be important to be held constant to give other variables a more accurate coefficient estimate. In the planning phase of this

research, a model based on linear regression instead of the linear probability model was anticipated. However, finding an outcome that was continuous and normally distributed became a problem. The continuous outcome that was intended to be used was days to new citation within a year period. The results show that the largest amount of juveniles reoffended within the first few days of release which skewed the outcome variable. Once a continuous outcome variable could not be found a dichotomous variable was implemented creating a linear probability model. This is based a similar model used in the research by Gottfredson and Snyder (2005) covered in the literature section above.

Discussion

While partial support was found for the two research hypotheses, it was immediately apparent that rescoring the RAI is not going to be enough to produce an accurate instrument. The best way to increase the accuracy of the RAI would be a complete restructuring of the tool by using more appropriate risk factors and recreating thresholds.

As discussed above, the Montana Risk assessment instrument was created by modifying Virginia's Detention assessment instrument which was created using a "consensus design." There are a number of reasons to believe that the DAI may not be the most effective model for a state like Montana.

To begin, the Virginia instrument was developed for a much larger urban population of people where the largest minority consisted of African Americans. In contrast, the counties that use the RAI in Montana have much smaller rural populations where the largest minority population is American Indian. While there were some changes made in the adoption of the RAI, these were insufficient to provide enough of the risk factors to adequately encompass the

Montana juvenile population. Moreover, the RAI kept the exact same thresholds that Virginia had ignoring the fact that on Montana's instrument juveniles can score 19 points more than on the DAI. If Montana was to create its own unique instrument the question still remains, could one tool be used in all 56 counties or is there a need for site specific tools? In the current analysis this problem is seen most clearly when comparing Hill County with a population of roughly 16,632 people (U.S. Census Bureau 2009) and a high proportion of American Indian juveniles with Yellowstone County with a population of roughly 144,797 people (U.S. Census Bureau 2009). Before site specific tools can be created a method for creating an accurate tool must be investigated.

The consensus method was designed for developing site specific instruments. It uses local stakeholders and professionals to choose the most appropriate risk factors for the area it will be used in. The flaw is that without running statistical tests on these risk factors it will be unknown if all or any of these factors are appropriate. Based on the results in this research using both logistic regression and the linear probability model, six of the nineteen risk factors were found to be negatively correlated with the outcome variable. This suggests that not only were these factors not significant predictors, but having them in the model results in a reduction in the likelihood of a citation in the period of risk. Moreover, only four risk factors were even close to being statistically significant.

The risk factor, "The youth was taken into custody on a valid warrant or pick up order" is an interesting risk factor that should be discussed. If a juvenile is brought into detention on a valid warrant or pickup order, the juvenile will receive 15 points on the RAI. This automatically places the juvenile into the "Secure Detention" indicated decision for the RAI. This risk factor is the highest weighted risk factor (tied with "Most serious offense alleged in current referral,

felonies against persons). However, no evidence was found to support the weight for this risk factor. It was the least significant of all of the risk factors in the linear probability model, and was found to be negatively correlated to the outcome variable in the logistic regression model. Further investigation on this risk factor is necessary to determine if such a large weight is appropriate or if the risk factor is even necessary in predicting if a juvenile will receive a new citation in the period of risk.

The RAI risk factors can be placed into two broad categories: Current offense/s, and history of offenses. These risk factors may be appropriate, however, they are not all encompassing of factors that predict if the juvenile will receive a new citation. A few examples of other factors that may help create a more predictive RAI are: the family and living situation of the juvenile, the friends of the juvenile, how well the juvenile is doing in school, drug use, and psychological issues in the juvenile. Risk factors found on the general recidivism risk assessment instrument "Level of Service Inventory-Revised" (LSI-R) that could potentially improve the RAI include, but are not limited to: Family employment, family income, recreation, alcohol, emotional stability and general attitude (Andrews and Bonta 1995). Finding risk factors on instruments that have already been created and have been proven to be predictive should be a starting point when looking for alternative risk factors. How to add these factors into the RAI without it becoming racially or gender biased will be another issue future research must consider.

Once the most appropriate risk factors have been uncovered it will still be important to examine thresholds for detention decisions. There will never be an exact formula that will predict which juveniles will receive a new citation and which juveniles will not. This is why when creating a RAI it is important to consider the errors when evaluating the thresholds. Pushing the thresholds into the higher scores will increase the false negative predictions. False negative

36

predictions occur when juveniles are predicted not to reoffend but do. Bringing the thresholds closer to the lower scores will increase the false positive predictions. False positive predictions occur when juveniles are predicted to reoffend but don't. Which error the RAI will make must be based on the policies that surround the RAI. Areas where community protection is the highest priority will lean toward false positive predictions while those communities where the juvenile's rights are the highest priority will learn toward false negative predictions.

Having discussed the issues that surround the Montana RAI, it is important to recognize that many states do not have any form of risk assessment tools. The RAI is a step in the right direction. It will be a slow process fine tuning the RAI and determining what method and which risk factors are most appropriate.

Future research must focus on three major topics: risk factors, weights, and cut points. Finding appropriate risk factors for the risk assessment in Montana should be the first priority. Adding appropriate weights and finding the most effective cut points cannot be investigated until the risk factors have been found. Risk factors need to reflect more than just the current offense and offense history of the juvenile. Once appropriate risk factors have been found, research similar to this will be necessary to determine which model is the most predictive and what cut points eliminate the most false positive and false negative predictions.

The current research was done using a retrospective research process in which the sample consisted of juveniles that have already been in detention and already had the opportunity to reoffend. It would be highly beneficial in the future to alter the current RAI and use a prospective approach in which a new RAI could be administered to juveniles as they come into the detention center before they are detained.

Research toward the most predictive RAI is imperative for the safety of Montana communities and the human rights of the juveniles that come into the justice system. It is expected that the results presented here provide and important step in a positive direction and that the findings and conclusions stimulate further inquiry into this important and timely issue.

References

Aguinis, Herman and Ryan K. Gottfredson. 2010. "Best-practice Recommendations for Estimating Interaction Effects Using Moderated Multiple Regression." *Journal of Organized Behavior* 31:776-786.

Andrews. D. A., and James Bonta. 1995. "The Level of Service Inventory-Revised: Color Plot Profile Form for Male Inmates. Toronto, Canada: Multi-Health Systems.

Anne E. Casey Foundation. Retrieved from the World Wide Web. www.aecf.org. March 14th 2012.

Burgess, E. W. 1928. Factors Determining Success or Failure on Parole. In A. A. Bruce, *The workings on the indeterminate sentence law and the parole system in Illinois.* Springfield: Illinois State Board of Parole.

Caulkins, Jonathan, Jacqueline Cohen, Wilpen Gorr and Jifa Wei. 1996. "Predicting Criminal Recidivism: A Comparison of Neural Network Models with Statistical Methods." *Journal of Criminal Justice* 24:227-240.

Gottfredson, Don M. and Howard N. Snyder. 2005. "The Mathematics of Risk Classification: Changing Data into Valid Instruments for Juvenile Courts." *Office of Juvenile Justice and Delinquency Prevention* 1-26.

Gottfredson, Stephen D. and Don M. Gottfredson. 1980. "Screening for Risk: A comparison of Methods." *Criminal Justice and Behavior* 7:315-330.

Hanson, Karl R. 1997. "The development of a brief actuarial risk scale for sexual offense

     recidivism." (User Reports 97-04). Ottawa: Department of the Solicitor General of

     Canada.

Hare. Robert D. 1991. "The Hare Psychopathy Checklist-Revised (PCL-R). Toronto Canada:

     Multi-Health Systems.

Harris, Grant, Marnie Rice, and Vernon Quinsey. 1993. "Violent recidivism of mentally

     disordered offenders: The development of a statistical prediction instrument." *Criminal*

     Justice and Behavior, 20:315-335

Hollist, Dusten, Jacob Coolidge, Wes Delano, Ian Greenwood, Mike King, Tyson McLean,

     Patrick Mckay, James Burfeind, Chuck Harris, Dan Doyle. Forthcoming. "The Montana

     Pre-Adjudicatory Detention Risk Assessment Instrument: A Validation and Assessment

     Study."

Kirby Bernard C. 2007. "Parole Prediction Using Multiple Correlation." *American Journal of*

     Sociology 59:539-550.

Loza, Wagdy, and Katherine Green. 2003. "The Self-Appraisal Questionnaire: A self-report

     measure for predicting recidivism versus clinician-administered measures: a 5-year

     follow-up study. *Journal of interpersonal Violence.* 18:781-797.

Reiner, Scott, Jared Miller and Tripti Gangal. 2006. "Public Safety Outcomes of Virginia's

     Detention Assessment Instrument." *Virginia Department of Juvenile Justice.*

Silver, Eric and Lynette Chow-Martin. 2002. "A Multiple Models Approach to Assessing

  Recidivism Risk: Implications for Judicial Decision Making." *Criminal Justice and*

  Behavior 29:538-568.

Silver, Eric, William R. Smith and Steven Banks. 2000. "Constructing Actuarial Devices for

  Predicting Recidivism: A Comparison of Methods." *Criminal Justice and Behavior*

  27:733-764.

Simon, Frances H. 1972. "Statistical Methods of Making Prediction Instruments." *Journal of*

  Research in Crime and Delinquency 9:46-53.

Steinhart, David. 2006. "Juvenile Detention Risk Assessment: A Practice Guild to Juvenile

  Detention Reform." *Juvenile Detention Alternatives Initiative: A Project of the Annie E.*

  Casey Foundation.

Thomas, Stuart, Morven Leese, Elizabeth Walsh, Paul McCrone, Paul Moran, Tom Burns,

  Francis Creed, Peter Tyrer, Thomas Fahy. 2004. "A Comparison of Statistical Models in

  Predicting Violence in Psychotic Illness." *Department of Forensic Mental Health*

  Science.

U.S. Census Bureau. 2009. American Fact Finder. Retrieved from the World Wide Web.

  http://factfinder2.census.gov. April 19th 2012.

Appendix A

DETENTION RISK ASSESSMENT INSTRUMENT – JPO

Youth's Name: _____ DOB: _____/_____/_____ Date: _____/_____/_____

Officer Completing Assessment: _____ Race: _____ Gender: M / F

Does youth meet statutory criteria for detention?	_____ No _____ Yes

If the youth does not fall into the numbered category please respond with a -0- for the categories point totals.

1. The youth was taken into custody on a valid warrant or pick up order..............................15   [+]

2. Most Serious Offense Alleged in Current Referral
Felonies against Persons. 15
Other Felonies. 10   [+]
Misdemeanors against Persons. 7
Other Misdemeanors. 5

3. Additional Offenses Alleged in Current Referral
One or More Additional Current Felony Offenses 5   [+]
One or More Additional Misdemeanor or Violation of Probation/Parole Offenses 3

4. Prior Admissions of Guilt
Two or more prior admissions of guilt for felony offenses.......................6
One prior admission of guilt for a felony offense....................................4
Two or more prior admissions of guilt for misdemeanor or status offenses....................3   [+]
Two or more prior admissions of guilt for probation/parole violations.....................2
1 prior admission of guilt for any misdemeanor or status 1

5. Referrals Pending Adjudication
One or more pending referrals for a felony offense.................................8
Two or more pending referrals for other offenses...................................5   [+]
One pending referral for other offense/offenses 2

6. Supervision Status
Intensive or Close Supervision (Drug/Treatment Court, House Arrest, Group Home, Etc)........10
Formal Release Conditions..5   [+]
On Probation...5
On Parole..5

7. Warrant History
Two or More Warrants..3
One Warrant..1   [+]

Total Score ....   [   ]

Indicated Decision: _____ 0-9 Release _____ 10-14 Detention Alternative _____ 15+ Secure Detention

Final Decision: Detain   Release   Release with conditions

Override Justification:

_____

_____

_____

_____

                    Override Approved: _____ Date: _____

Probation Officer: _____ Date: _____ Time: _____