
Predictive Validity of the COMPAS Reentry Risk Scales

AN OUTCOMES STUDY CONDUCTED FOR
THE MICHIGAN DEPARTMENT OF CORRECTIONS:
UPDATED RESULTS ON AN EXPANDED RELEASE SAMPLE

WILLIAM DIETERICH, PhD
WILLIAM OLIVER, PhD
TIM BRENNAN, PhD

AUGUST 22, 2013



Contents

Table of Contents	i
1 Executive Summary	1
2 Materials and Methods	6
2.1 Sample	6
2.2 Measures	11
2.3 Analytic Approach	14
3 Results	19
3.1 Description of Events of Interest During the Follow-up . . .	19
3.2 Predictive Accuracy Results	23
3.2.1 AUC Summary	23
3.2.2 ROC Methods Summary	28
3.2.3 Failure Probabilities Across the Risk Scale Levels . . .	32
3.2.4 Survival Plots	36
A 2012 Reentry Study AUC Results	40
Bibliography	43

Chapter 1

Executive Summary

This report presents results from a prison reentry outcomes study conducted for the Michigan Department of Corrections (MDOC). The report includes updated results on the 2008-2010 release sample that was examined in the prior COMPAS Reentry outcomes study conducted for the MDOC in 2012 (Dieterich, Oliver, & Brennan, 2012). The study sample has been expanded beyond the 2008-2010 release sample to include releases from 2008 through 2012.

Objectives of this report: The main objective of the current study was to test the predictive validity of the COMPAS Reentry default risk scales in the 2008-2012 release sample with extended follow-up. A secondary objective of the study was to construct the data sets required for other research projects being conducted for the MDOC including the development, testing, and calibration of the new VFO and Non-VFO Reentry Risk Scales. Those research projects are covered in separate reports (Northpointe Inc., 2013; Oliver, Dieterich, & Brennan, 2013b, 2013a).

Study design: There were 53,131 unique individuals released from June 2008 through December 2012. There were 48,223 releases matched with a COMPAS Core or Reentry assessment. There were 47,284 matched individuals assessed with COMPAS within 730 days prior to their release. There are 45,809 individuals with complete data on the risk scales and risk scale inputs representing the overall study sample. In the overall study sample 85% of the assessments were conducted with COMPAS Reentry. Approximately 95% of the study sample was released onto parole.

In the current study we test the predictive validity of the General Recidivism Risk Scale with the following outcomes: any offense arrest, felony offense arrest, abscond, and new commitment. The validity of the Violent Recidivism Risk Scale is tested with the following outcomes: Violent Felony Offense (VFO)

arrest and any violent offense arrest (misdemeanor or felony). Predictive validity is demonstrated by the ability of the risk scales to accurately discriminate recidivists from non-recidivists for the study outcomes. In the current study we use receiver operating characteristic (ROC) methods to estimate the area under the ROC curve (AUC) to measure discrimination. Following conventions in the field of criminology, we define AUCs of .70 and higher as “good.” The AUC is described in more detail in the methods section.

Major findings: The findings in the current study with extended follow-up support the previous findings of good overall accuracy for the General Recidivism Risk and Violent Recidivism Risk scales in predicting the absconding and new commitment outcomes.¹ The General Recidivism Risk and Violent Recidivism Risk scales also demonstrated good accuracy for predicting the arrest outcomes. As in previous studies the findings suggest that the risk models perform better for *first releases*, but the differences in accuracy among *first releases* versus all release are smaller than the differences observed in previous studies. For the arrest outcomes only small differences in the performance of the risk scales for *first releases* versus *all releases* were observed.

As in the previous study, the General Recidivism Risk levels and the Violent Recidivism Risk levels effectively divided cases into low-, medium-, and high-risk groups with distinct base rates of failure.

1. *Predictive accuracy of the risk scales:* The COMPAS Reentry General Recidivism Risk and Violent Recidivism Risk scales were evaluated for their utility in predicting any offense arrest, felony offense arrest, violent offense arrest, VFO arrest, absconding, and new commitment. Results from logistic regression models and survival models provide good support for the predictive validity of the General Recidivism Risk Scale. Using a standard approach for evaluating predictive accuracy in logistic regression models and survival models, we found that the General Recidivism Risk decile score has good predictive accuracy for any offense arrest, felony offense arrest, and new commitment. Results from logistic regression models and survival models indicate that the Violent Recidivism Risk decile score has good accuracy for predicting violent offense arrest and VFO arrest.

As we found in the 2012 outcomes study, the General Recidivism Risk scale demonstrated good accuracy for predicting absconding. The results from logistic regression models were similar to the results obtained in the previous study.

¹In this report we use the term accuracy to refer to the ability of the risk scale to accurately discriminate recidivists from non-recidivists.

The extended follow-up allowed us to examine three-year outcomes in estimation samples of substantially increased size over the previous study. Estimation samples of 16,019, 14,473, and 10,533 released inmates with at least three years of follow-up on the street were obtained for the Felony, VFO, and New Commitment outcomes, respectively.

Tables 3.4 through 3.7 show the AUC results for the abscond and new commitment outcomes. Tables 3.4 and 3.6 summarize the AUC results from the logistic regression models for *first releases* (parole and maxout) who had at least two years and three years of follow-up time, respectively. The results for *all releases* with two years and three years of follow-up are reported in Tables 3.5 and 3.7, respectively. The AUC results from the 2012 study are shown in Appendix A.

Tables 3.8 through 3.11 show the AUC results for the arrest outcomes. Tables 3.8 and 3.10 show the AUC results from the logistic regression models for *first releases* who had at least two years and three years of follow-up time, respectively. The results for all releases are reported in Tables 3.9 and 3.11.

The results indicate that the risk scales accurately discriminate the study outcomes. Importantly, the ability of the risk scales to discriminate the different outcomes holds up when the follow-up is extended to three years.

The AUCs from logistic regression models fitted to *all releases* with at least two years of follow-up for felony arrest (.71), VFO arrest (.72), and new commitment (.70) were very similar to the AUCs for felony arrest (.71), VFO arrest (.71), and new commitment (.69) obtained from models fitted to the subset with at least three years of follow-up.

As was the case in prior outcomes studies conducted for the MDOC (Brennan & Dieterich, 2008; Brennan, Dieterich, & Khan, 2009; Dieterich, Brennan, & Breitenbach, 2010; Dieterich, Oliver, & Brennan, 2011; Dieterich et al., 2012), we found that when we controlled for type of commitment, the accuracy of the risk scales increased. A comparison of the AUCs in the tables shows that, for all outcomes and models, higher AUCs were obtained when making predictions among *first releases*. This subset excludes parole technical violators and parole violators with a new sentence.

The differences in accuracy results obtained among *first releases* versus *all releases* were less pronounced in the current study in comparison to previous studies. For the arrest outcomes only small differences in the performance of the risk scales for *first releases* versus *all releases* were observed.

2. *Percentage Recidivating Across the Risk Scale Levels:* Results demonstrate that the Recidivism Risk Scale levels (Low, Medium, High) effectively resolve the reentry outcomes sample into groups with distinct recidivism probabilities. Findings show that inmates scoring at the high level of the General Recidivism Risk Scale have an odds of arrest within three years that is approximately six times the odds of those scoring at the low level of the scale. Results from models predicting a new commitment indicate that inmates scoring at the high level of the General Recidivism Risk Scale have an odds of new commitment within three years of release that is approximately four times the odds of those scoring at the low level of the scale.

Results demonstrate that the Violent Recidivism Risk Scale levels (Low, Medium, High) effectively resolve the reentry outcomes sample into groups with distinct recidivism probabilities. Findings show that the odds of a VFO arrest within three years of release for persons in the high-risk level are approximately five times the odds of persons in the low-risk level.

3. *Comparison with Previous MDOC Outcomes Results:* The AUCs obtained in the previous study are reported in Tables A.1 through A.4 in the appendix. Comparing the logistic regression results for *all releases*, the AUC of .73 obtained for absconding in the current study (2-year follow-up) is somewhat higher than the AUC of .70 obtained in the 2012 study (1.5-year follow-up). The AUC for felony arrest in the current study (.71) is the same as the result obtained in the 2012 study (.71). The AUC for VFO arrest in the current study (.72) is marginally higher than the AUC obtained in the 2012 study (.71). The AUC for new commitment in the current study (.70) is similar to the AUC for new commitment obtained in the 2012 outcomes study (.69).

Conclusion: The findings confirm the results of the previous studies that demonstrated the predictive validity of the default risk scales. The current study goes further because it included an expanded sample and extended follow-up. The updated follow-up made it possible to test the predictive validity of the risk scales at three years post-release. The three year follow-up is an important milestone because the MDOC defines recidivism as a return to prison within three years.

Next Steps: The outcomes data sets that were constructed for the current study were also used as training and test data sets to develop new risk models for reentry with improved performance with parole violators returned to prison for a technical violation or new sentence. These outcomes data sets were also used for testing and calibration of probability versions of the new risk scales to replace the current decile cutting points scheme in use in the COMPAS Reentry

application in the MDOC ([Oliver et al., 2013b](#)). Continued periodic follow-ups on the 2008-2012 release sample are recommended to support ongoing development, testing, and calibration.

Chapter 2

Materials and Methods

2.1 Sample

The MDOC reentry sample consists of all persons who were released from prison from June 2008 through December 2012 and assessed with COMPAS within 730 days prior to release. There were 53,131 unique individuals released from June 2008 through December 2012. There were 48,223 releases matched with a COMPAS Core or Reentry assessment. There were 47,284 matched individuals assessed with COMPAS within 730 days prior to their release.¹ There are 45,809 individuals with complete data on the risk scales and risk scale inputs representing the overall study sample.² In the overall study sample 85% of the assessments were conducted with COMPAS Reentry. Table 2.1 reports the number of individuals included in the Reentry study sample from each release facility.

In the overall sample there were 11,717 (25.6%) cases released into the follow-up after a prior release and return for a parole violation.³ These cases include 4,700 parole technical violators (PVTECH) and 7,017 parole violators with a new sentence (PVNS). They enter the study by either being re-released back onto parole or discharged on the maximum after parole and return.

Depending on the analysis, two types of samples are used: (1) The overall study sample of 45,809 released inmates is used for survival analysis, and (2) Subsamples of inmates with at least three years of follow-up on the street are used for approaches that require binary outcomes, including logistic regression

¹887 out-of-state parole transfers were dropped from the current analyses.

²198 cases were dropped because the MDOC and SID could not be matched with MSP data.

³145 cases in the overall sample were missing the intake movement code in the transit file.

and ROC methods. In the current report we refer to these two types as survival and logistic regression samples, respectively. The sample sizes for the approaches that use binary outcomes ranged from 10,532 for the new commitment outcome to 16,245 for the any arrest outcome. Different sample sizes result from failure-specific patterns of competing events. For example there are more events competing with the VFO outcome ($n=14,475$) than with the any arrest outcome. Another example is that the follow-up end date on the commitment outcome is backdated, resulting in fewer cases that have a minimum of three years on the street for this outcome.

Table 2.1: Number of inmates in the overall sample grouped by the facility from which they were discharged prior to entering the study.

Facility	Frequency	Percent
Alger Maximum Facility	4	0.0
Baraga Maximum Facility	15	0.0
Bellamy Creek Correctional Facility	2788	6.1
Boyer Road Correctional Facility	59	0.1
Camp Branch	52	0.1
Camp Cassidy Lake Men	2618	5.7
Camp Cassidy Lake Women	393	0.9
Camp Cusino	2	0.0
Camp Lehman	65	0.1
Camp Ottawa	1	0.0
Camp Valley	44	0.1
Camp White Lake	32	0.1
Carson City Correctional Facility	830	1.8
Charles Egeler Reception Center	538	1.2
Chippewa Correctional Facility	88	0.2
Clare County Jail	40	0.1
Cooper Street Facility	4041	8.8
Deerfield Correctional Facility	67	0.1
Earnest C. Brooks Correctional Facility	274	0.6
Florence Crane Correctional Facility	425	0.9
G. Robert Cotton Correctional Facility	2058	4.5
Gus Harrison Correctional Facility	2884	6.3
Hiawatha Correctional Facility	30	0.1
Huron Valley Complex/Mens	11	0.0
Huron Valley Complex/Womens	2320	5.1
Ionia Maximum Correctional Facility	207	0.5
Jackson County Jail	27	0.1
Kinross Correctional Facility	134	0.3

Table 2.1: (continued)

Facility	Frequency	Percent
Lakeland Correctional Facility	311	0.7
Lenawee County Jail	8	0.0
Macomb Correctional Facility	3281	7.2
Marquette Branch Prison	188	0.4
Mason County Jail	3	0.0
Men's Electronic Monitoring Center	1596	3.5
Michigan Reformatory	236	0.5
Mid-Michigan Correctional Facility	1315	2.9
Mound Correctional Facility	3013	6.6
Muskegon Correctional Facility	59	0.1
Newberry Correctional Facility	91	0.2
Oaks Correctional Facility	175	0.4
Ojibway Correctional Facility	21	0.0
Osceola County Jail	19	0.0
Parnall Correctional Facility	4725	10.3
Parole Office	1	0.0
Parr Highway Correctional Facility	119	0.3
Pine River Correctional Facility	230	0.5
Pugsley Correctional Facility	1401	3.1
Richard A Handlon Correctional Facility	447	1.0
Robert Scott Correctional Facility	52	0.1
Ryan Correctional Facility	2911	6.4
Saginaw Correctional Facility	2598	5.7
St. Louis Correctional Facility	102	0.2
Standish Maximum Corr. Facility	47	0.1
Straits Correctional Facility	13	0.0
Thumb Correctional Facility	264	0.6
VanBuren County Jail	11	0.0
West Shoreline Correctional Facility	2496	5.4
Women's Electronic Monitoring Center	3	0.0
Woodland Center Correctional Facility	26	0.1
Sum	45809	100.0

The outcomes sample consists of 45,809 individuals with complete data on the General Recidivism Risk and Violent Recidivism Risk scales. In the outcomes sample, 94% of the offenders are male ($n = 42,938$) and 6% are female ($n = 2,871$). The age of the offenders in the estimation sample ranges from 16.3 to 89.2 years. The average age in the sample is 36.5 years (valid $N = 45,809$). The ethnic composition of the sample is 48% White, 47% African American,

Table 2.2: Distribution of top commitment offense categorized by MDOC offense groups in the MDOC reentry study sample.

offense group	frequency	relative frequency
homicide	1489	0.03
robbery	3870	0.08
criminal sexual contact	6493	0.14
assault	5189	0.11
arson	307	0.01
other sex offenses	335	0.01
assaultive other	3789	0.08
burglary	4638	0.10
larceny	2519	0.06
fraud	1408	0.03
forgery	1425	0.03
motor vehicle	516	0.01
malicious destruction	131	0.00
weapons possession	3114	0.07
drug	6727	0.15
ouil 3rd	2315	0.05
other non-assaultive	1542	0.03
missing	2	0.00
Sum	45809	1.00

and 5% Other.

Table 2.2 shows the distribution of the current commitment offense groups in the reentry sample. Note that we use the offense groups taken from the MDOC information management system (IMS), as opposed to the offense groups as collected on the COMPAS questionnaire. This is done because there are discrepancies between the two data sources in how commitment offenses were identified and categorized. To maintain consistency throughout the report we use the MDOC MIS categorizations.

Collapsing further, we find that 38% of the offenses are categorized as nonassaultive, 15% as drug related, 14% as criminal sexual contact, and 33% as other assaultive. The frequency distributions in the reentry sample for commitment type and and discharge type are shown in Tables 2.3 and 2.4, respectively.

Table 2.3: Distribution of commitment type in the MDOC reentry study sample.

commitment type	frequency	relative frequency
New Commitment, Probation Technical Violator	5209	0.11
New Commitment, Probationer w/ New Sentence	6848	0.15
New Commitment	21816	0.48
Parole Violator w/ New Sentence	7017	0.15
Additional Sentence Imposed	74	0.00
Returned as Technical Parole Violator	4700	0.10
Missing	145	0.00
Sum	45809	1.00

Table 2.4: Distribution of discharge type in the MDOC reentry study sample.

discharge type	frequency	relative frequency
Discharged on Maximum After Parole and Return	189	0.00
Discharged on Maximum w/o Parole	3081	0.07
Reinstated on Parole from Institution	672	0.02
First Parole this Term/Sentence	39241	0.86
Paroled in Custody	1417	0.03
Reparole While on Same Term	1209	0.03
Sum	45809	1.00

Note. First Parole this Term/Sentence group includes some PVTECH ($n=3,122$) and PVNS ($n=6,445$).

2.2 Measures

Independent Variables

The General Recidivism Risk and Violent Recidivism Risk scales are the central independent variables in the study. Additional independent variables include the criminal history scales and the needs scales in Reentry COMPAS. The risk scales, criminal history scales, and needs scales are described in the Reentry COMPAS documentation. Table 2.5 shows the summary statistics and alpha coefficients for the Reentry COMPAS scales in the study sample.

There are eleven scales that have alpha coefficients below .70. Low alphas indicate the scale has too few items or the items don't have much in common and possibly measure more than one construct (Nunnally & Bernstein, 1994). The low alphas on Violence History (.51), Current Violence (.50), Prison Misconduct (.61), and Gang Affiliation (.40) are not unexpected because these scales are indexes of different types of offenses or behaviors. A low alpha does not indicate a problem because the items are not expected to be highly correlated. Family Crime (.62) is a similar type of index of problems experienced by family members.

Table 2.5: Summary of the Reentry COMPAS scales in the outcomes sample.

	Items	N	Min	Max	Mean	SD	Alpha
Early Onset	4	38278	0.00	6.00	1.14	1.73	0.80
Criminal Involvement	4	38976	0.00	19.00	7.61	4.66	0.76
Noncompliance History	5	38976	0.00	21.00	4.39	3.99	0.68
Violence History	9	38976	0.00	19.00	2.03	2.27	0.51
Current Violence	7	38761	7.00	13.00	8.32	1.27	0.50
Prison Misconduct	7	38282	0.00	10.00	1.29	1.67	0.61
Gang Affiliation	3	38289	0.00	6.00	0.11	0.50	0.40
Substance Abuse	6	38289	0.00	12.00	6.59	3.92	0.78
Housing Problems	4	38288	0.00	8.00	1.26	1.92	0.75
Financial Problems	6	38289	0.00	13.00	3.65	2.73	0.68
Social Environment	5	37972	0.00	12.00	3.38	2.91	0.80
Reentry Employment Problems	12	38289	0.00	21.00	6.05	4.31	0.78
Voced Problems	11	38976	11.00	30.00	19.16	3.95	0.71
Low Family Support	5	38289	0.00	6.00	1.90	1.62	0.68
Family Crime	6	38289	6.00	12.00	7.66	1.53	0.62
Social Isolation	5	38289	0.00	10.00	1.21	1.63	0.67
Purpose/Direction	7	38289	0.00	14.00	1.84	2.15	0.62
Low Empathy	5	38289	0.00	20.00	5.06	2.60	0.70
Depression	6	38273	0.00	6.00	0.48	1.17	0.82
Anger	7	38289	0.00	14.00	1.80	2.71	0.78
Criminal Thinking Obs.	7	38289	0.00	14.00	2.27	3.45	0.90
Criminal Thinking SR	10	38289	13.00	59.00	29.51	5.89	0.83
Criminal Personality	13	38289	0.00	42.00	8.29	7.08	0.67
Negative Cognitions	21	38287	10.00	50.00	18.84	5.41	0.85
Low Efficacy	15	38289	0.00	30.00	2.83	3.64	0.84
Violence Recidivism Risk	25	38976	-4.82	2.01	-1.83	0.97	NA
General Recidivism Risk	25	38976	-3.26	3.73	-0.25	0.80	NA

Cronbach's reliability coefficient alpha: .70 and above=good internal consistency.

Alphas not available for risk scales because they are linear equations. Table does not include the COMPAS Core subsample.

Outcomes Measures

We examine the utility of the General Recidivism Risk and Violent Recidivism Risk scales for predicting the outcomes any arrest, felony arrest, violent arrest, VFO arrest, absconding, and new commitment. The following provides definitions for the outcomes of interest in the study:

- **Any Offense Arrest:** The first arrest for any misdemeanor or felony offense following release. Arrests were determined through computer match with the Michigan State Police (MSP) criminal history records. MDOC Research submitted a request to the MSP to conduct a search of Criminal History Records for all inmates released since January 2008. For the 46,007 inmates in our overall sample, 45,809 were matched on MDOC number (PRN) or State Identification Number (SID), and 198 were not matched. The MSP file includes incident, arrest, charge, and judicial records. The incident records are linked with arrest, charge, and judicial records by a criminal tracking number (CTN) assigned by the prosecutor. We defined an arrest as the presence of an incident record with an accompanying arrest record. We matched 28,807 criminal history records with our study sample that had an incident date after the prison release date with an accompanying arrest record, involving 16,053 individuals. We included only arrest records with a Criminal Tracking Number assigned by the prosecutor.

The *incident date* in the MSP Criminal History Record was used to determine the failure point. The MSP data were extracted on May 28, 2013. To account for data entry lag time, the analysis cut-off date for the arrest outcomes was set back to April 1, 2013. The arrest follow-up period ends on the MSP back date, death, or return to prison, whichever occurred first.

- **Violent Offense Arrest:** The first arrest for a violent offense following release. This outcome was determined through the MSP Criminal History Records match. An arrest was determined using the same methods that were used for the arrest outcome described above. The MSP file includes incident, arrest, charge, and judicial records. The charge and judicial records include an MCL statute code that was matched with the MCL crosswalk. Violent offenses were determined using the crosswalk of Michigan Compiled Laws (MCL) provided by MDOC Research.
- **Felony Offense Arrest:** The first arrest for a felony offense following release. This outcome was determined through the MSP Criminal History Records match. Felony arrest was determined using the same methods that were used for the arrest outcome described above. The level of the

offense (misdemeanor or felony) was determined with the Prosecuting Attorneys Coordinating Council (PACC) crosswalk or the arrest type on the MSP arrest record. The arrest type (on the MSP arrest record) was concordant with the PACC code on approximately 99% of the records. We treated high misdemeanors (H) as misdemeanors in the analysis.

- **VFO Arrest:** The first arrest for a violent felony offense following release. This outcome was determined through the MSP Criminal History Records match. A violent felony arrest was determined using the same methods that were used for the violent arrest outcome described above. Violent offenses were determined using the crosswalk of Michigan Compiled Laws (MCL) provided by MDOC Research as described above. The level of the offense (misdemeanor or felony) was determined with the Prosecuting Attorneys Coordinating Council (PACC) crosswalk or the arrest type category on the MSP arrest record.
- **Abscond:** The first abscond observed in the movement file following release from prison. The movement date determines the failure point. The follow-up ends on the date the movement file was created (June 6, 2013) or in the event of death, parole discharge, or return to prison, whichever occurred first.
- **New Commitment:** The first arrest for a felony following release that resulted in a new commitment. The felony offense date associated with the new commitment determines the failure point. We use the code for new commitment from the movement file and offense dates from the sentencing file to define and determine this outcome. To allow time for disposition and sentencing the follow-up cutoff was backdated eight months from the movement file extract date (October 6, 2012). The follow-up ends on the commitment backdate or in the event of death or return to prison for a technical parole violation, whichever occurred first.

2.3 Analytic Approach

Survival Models

We fit survival models to the study outcomes misdemeanor offense arrest, felony offense arrest, VFO arrest, absconding, and new commitment. Survival models are appropriate for these data because we are interested in both the occurrence and timing of the outcomes. Survival time begins on the release from prison date. The risk set at this point contains all inmates in the estimation sample (valid $N = 45,809$). Survival time is measured in days from

release date to the point of the first failure of interest, first competing event, or end of the follow-up, whichever occurs first. For new commitment the failure time point is determined by the felony offense date in the sentencing file. For absconding the failure time point is determined from the movement file. For the arrest outcomes the failure time point is determined by the incident date in the MSP criminal history record. Cases that do not fail by the event of interest date or a competing event during the follow-up are censored at the end of the study. For absconding the end of study is determined by the date that the movement file and sentencing file were created (June 6, 2013). For new commitment offense, the end of study is determined by the commitment back date (October 06, 2012). For the arrest outcomes, the MSP back date (April 1, 2013) is the end of study. Cases remain in the risk set and contribute information to the analysis until the point of failure of interest, occurrence of a competing event, or the end of study, whichever occurs first.

Cases can fail by an event of interest such as a new felony arrest as well as the competing events new commitment, return for a technical violation, discharge from parole, and death. These competing events alter the probability of observing the event of interest. For example, in the felony arrest model, if a parolee is returned to prison for a parole technical violation before we observe a felony offense arrest, they cannot fail by the event of interest (i.e., a new felony arrest) - at least while in prison for the competing reason. Even though a person could reoffend after discharge or after return and release back onto parole, we only tracked cases up to the first occurrence of either the event of interest or the competing event. Thus, within our survival models new commitment, return for a technical violation, discharge from parole (abscond model only), and death are treated as terminal or absorbing competing events.

- Any Arrest competing events: returns for technical violations, new commitments, and death.
- Felony Arrest competing events: returns for technical violations, new commitments, and death.
- Violent Arrest competing events: returns for technical violations, new commitments, and death.
- VFO Arrest competing events: returns for technical violations, new commitments, and death.
- Abscond competing events: returns for technical violations, new commitments, parole discharge, and death.
- New commitment competing events: returns for technical violations and death.

Logistic Regression

In addition to survival models we fit logistic regression models to the outcomes. In these models we create a binary outcome coded 1 if the case failed by the event of interest within two years of release and coded 0 if the case did not fail within two years or experience a competing event. We used the analysis time from the survival models that we fitted to calculate the number of street days from release until failure. The survival models control for custody holds and competing events. From these models we obtained a measure of days on the street from release to the point of failure for each outcome we examined. For each event we used the results from the survival models to calculate time until failure and potential follow-up days. The logistic regression models predicting failure within two years from release excluded cases that did not have at least two years of potential follow-up time on the street. We defined three-year binary outcomes using this same approach.

Measures of Association

- *Failure Probability.* In typical survival data without competing events the Kaplan-Meier statistic is used as an estimate of survival or failure probability (1 - KM) at different time points. However, in the context of competing risks, 1 - KM is not a proper or interpretable failure probability. Competing events also violate the independence assumption that was discussed above in the context of informative censoring (e.g., see [Putter, Fiocco, & Geskus, 2007](#)). In this report we calculate the crude cumulative incidence function. This gives the probability of failure adjusted for failures from competing events. We also use a specialized proportional hazards survival model for competing risks ([Fine & Gray, 1999](#)) to estimate and plot failure probabilities (cumulative incidence curves) for each type of failure within the levels of the risk scores. The goal is to compare the **probability of the event of interest** (e.g., felony arrest) between the levels of the risk scores. The generality of the results from competing risk models is limited to populations with similar characteristics and similar patterns of competing events ([Pintilie, 2007](#)).
- *Odds Ratio.* We use logistic regression models to estimate the predicted probability of felony arrest, VFO arrest, absconding, and new commitment across the levels of the risk scales. We also calculate directly from simple cross tabulations the percentage of cases that fail (observed probability) across the risk scale deciles.
- *Area Under the Curve.* The overall predictive power of each risk scale is assessed with receiver operating characteristic analysis (ROC). The ROC

analysis is used to evaluate how accurately the risk scale discriminates between recidivists and nonrecidivists. The analysis produces a plot of the sensitivity against one minus specificity called the ROC curve. Sensitivity is the percentage of failures correctly classified and specificity is the percentage of successes correctly classified. One minus specificity is the false positive rate. Thus, the ROC curve is a plot of the true positive rate against the false positive rate for each cut-point of the scale. For the models that we fitted for this report, the cut-points are at the deciles of the risk scale scores. The area under the curve provides an overall measure of the performance of the risk scale. Larger areas under the curve indicate better predictive power. The area under the curve for a scale with no predictive power is .50. If the scale predicts perfectly, the area under the curve is 1. For binary outcomes, such as those in logistic regression models, the area under the curve can be interpreted as the probability that a randomly selected failure will have a higher risk score than a randomly selected success.

The AUC is a good method for evaluating prognostic models because the estimate is not influenced by the base rate (proportion of the sample that fails). This characteristic makes it easier to compare results across studies. However, the base rate can affect the precision of the AUC. The performance of any instrument will vary depending on the population (e.g., prison releases or probation cases), measurement error in the scale or outcome, and type of outcome (e.g., felony arrest versus abscond or new commitment).

The AUCs of the other main instruments often used for offender risk prediction may help to contextualize the findings in our study. Perhaps the best known instruments are the Violence Risk Appraisal Guide [VRAG] (Quinsey, Harris, Rice, & Cormier, 1998); the Level of Services Inventory-Revised [LSI-R] (Andrews & Bonta, 1995); and the Psychopathy Checklist-Revised [PCL-R] (Hare, 1991). The AUC values for these instruments in recent studies are quite varied according to the specific populations, outcome periods, and dependent variables used in specific studies.

(1) VRAG: Quinsey et al. (1998) found an AUC of 0.76 in a large scale, multiyear recidivism study. Barbaree, Seto, Langton, and Peacock (2001) reported AUCs of 0.69 in predicting serious re-offending and 0.77 when predicting any re-offense for sex offenders. Kroner, Stadtland, Eidt, and Nedopil (2007) obtained an AUC of 0.703 in a study of re-offending among mentally ill offenders.

(2) LSI-R: The recent review by Andrews, Bonta, and Wormith (2006) did not provide AUCs for the LSI-R. However, Barnoski and Aos (2003)

found AUCs of 0.64-0.66 for the LSI-R in predicting felony and violent recidivism among Washington State prisoners. Flores, Lowenkamp, Smith, and Latessa (2006) found an AUC of 0.689 using the LSI-R to predict reincarceration among federal probationers. Dahle (2006) reported an AUC of 0.65 using the LSI-R to predict violent recidivism. Barnoski (2006) reported an AUC of 0.65 using the LSI-R to predict felony sex recidivism.

(3) PCL-R: AUC levels again varied across studies. For example, in a Swedish study of mentally ill violent offenders, Graun, Belfrage, and Tengstrom (2000) found AUC levels of 0.64-0.75 based on various follow-up timeframes. Barbaree et al. (2001) reported AUCs of 0.61, 0.65, and 0.71 for the PCL-R in predicting various recidivism outcomes among sex offenders.

As more studies report AUCs for specific instruments, populations, outcome variables, and timeframes, it may become possible to identify which instruments perform well under these various conditions.

Chapter 3

Results

3.1 Description of Events of Interest During the Follow-up

Table 3.1 shows the number of events observed in each failure-specific survival model. The number failed gives the number of first failures for each event type. Some cases had multiple events of the same type, so the actual total number of events observed in the sample over the follow-up was higher than what is reflected in the table. Note that the arrest outcomes are tracked up to April 1, 2013 (MSP back date), but the new commitment outcome is only tracked up to October 6, 2012 (commitment back date). The abscond outcomes are tracked up to June 6, 2013 (transit file extract date). The difference in length of follow-up should be taken into account when comparing the base rates for the event types in the table.

Table 3.1: Distribution of Events for Each Failure-Specific Model

Failure Model	Censored	Failed	All
Any Arrest	31507	14302	45809
Felony Arrest	35073	10736	45809
Violent Arrest	41940	3869	45809
VFO Arrest	41596	4213	45809
Abscond	34640	8004	42644
Commitment	40760	5049	45809

Table 3.2 shows the distribution of time at risk from the survival models fitted to the study outcomes. Time at risk is the number of days from release to

first failure, competing event, or end of study. The competing events for each outcome model are described in the methods section.

Table 3.2: Description of time at risk from the survival model for each outcome

Outcome	Min	Median	Max
Any Arrest	1	543	1783
Felony Arrest	1	612	1783
Violent Arrest	1	652	1783
VFO Arrest	1	683	1783
Abscond	1	423	1779
New Commitment	1	564	1606

Figure 3.1 shows a plot of the crude cumulative incidence curves of first abscond for cases released onto parole in the presence of parole discharge and return to prison (returns and new commitments). The jumps in the discharge curve reflect inmates being discharged at one year, one and one half years, and two years. Here parole discharge and return to prison alters the probability of absconding. Both parole discharge and return to prison compete with absconding.

Table 3.3 shows the top felony charge (count) on the first felony arrest observed during the follow-up that was used as the outcome in the arrest models to test the predictive accuracy of the General Recidivism Risk and Violent Recidivism Risk scales. The felony counts on the first felony arrest event during the follow-up for each individual were ranked by the offense group variable from the MCL Crosswalk.

Table 3.3: Distribution of top felony charge (count) on the first felony arrest observed during the follow-up period.

offense group	frequency	relative frequency
homicide	96	0.01
robbery	629	0.05
criminal sexual contact	204	0.02
assault	2493	0.20
arson	20	0.00
other sex offenses	57	0.00
assaultive other	894	0.07
burglary	1058	0.09
larceny	1097	0.09
fraud	819	0.07
forgery	268	0.02
motor vehicle	155	0.01
malicious destruction	58	0.00
weapons possession	568	0.05
drugs	2623	0.22
ouil 3rd	298	0.02
other non-assaultive	820	0.07
Sum	12157	1.00

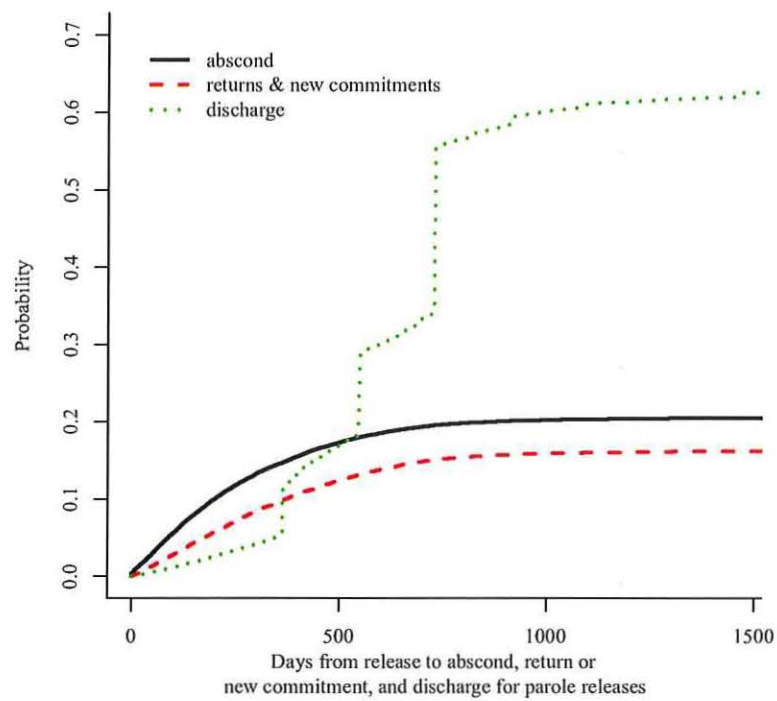


Figure 3.1: Crude cumulative incidence curves of absconding, new commitment, and parole discharge for all parole releases ($n=42,644$).

3.2 Predictive Accuracy Results

3.2.1 AUC Summary

Tables 3.4 through 3.7 show the AUC results obtained with the ROC method in the logistic regression samples for the abscond and new commitment outcomes. Tables 3.4 and 3.6 summarize the AUC results for *first releases* (parole and maxout) who had at least two years and three years of follow-up time, respectively. The results for *all releases* with two years and three years of follow-up are reported in Tables 3.5 and 3.7, respectively. The AUC results from the 2012 study are shown in Appendix A.

As was the case in prior outcomes studies conducted for the MDOC (Brennan & Dieterich, 2008; Brennan et al., 2009; Dieterich et al., 2010, 2011, 2012), the accuracy of the risk scales appears to be affected by commitment type. A comparison of the AUCs in Tables 3.4 through 3.7 shows that for all outcomes and models higher AUCs are obtained when making predictions among first releases. This subset excludes parole technical violators and parole violators with a new sentence.

Tables 3.8 through 3.11 show the AUC results for the arrest outcomes. Tables 3.8 and 3.10 show the AUC results obtained with the ROC method in the logistic regression samples for *first releases* who had at least two years and three years of follow-up time, respectively. The results for all releases are reported in Tables 3.9 and 3.11.

The differences in accuracy results obtained among first releases versus all releases were less pronounced in the current study in comparison to previous studies. For the arrest outcomes only small differences in the performance of the risk scales for first releases versus all releases are evident. For example, the AUC for the General Recidivism Risk Scale predicting any arrest within two years among first releases is .74 (Table 3.10). But the AUC for this model among all releases only drops to .73 (Table 3.11).

In general the *c* statistics obtained in the survival samples were lower in comparison to the AUCs obtained in the logistic regression samples (results not shown). The *c* statistics were typically about two points lower in the survival samples, compared with the AUCs in the logistic regression samples. The lowest *c* statistics were obtained in survival samples of all releases and in the abscond model. We speculate that the lower discrimination obtained in the survival samples is associated with the correlation of the risk scale with the competing events and a high base rate for the competing event.

Note that the number of women failing in some of the models was very small. The effective sample size for an analysis is determined by the number of fail-

ures. When the number of failures drops below 100, the AUC estimate lacks precision. The results for women on some outcomes, especially the VFO and new commitment outcomes, should be interpreted with caution. Attention should be given to the width of the confidence intervals of the AUC estimates.

Table 3.4: AUC results in logistic regression samples (first releases) with 2-year follow-up for felony arrest resulting in a new commitment and 1.5-year follow-up for the abscond outcome.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Abscond: Women	1142	222	0.73	0.69	0.76
Abscond: Men	14351	2046	0.73	0.72	0.74
Abscond: Total	15493	2268	0.73	0.72	0.74
Commitment: Women	1136	78	0.72	0.66	0.78
Commitment: Men	14440	1646	0.72	0.71	0.73
Commitment: Total	15576	1724	0.72	0.71	0.73

Note. Predicting abscond and new commitment with the General Recidivism Risk Scale.

Table 3.5: AUC results in logistic regression samples (all releases) with 2-year follow-up for arrest resulting in a new commitment and 1.5-year follow-up for the abscond outcome.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Abscond: Women	1394	338	0.71	0.67	0.74
Abscond: Men	19652	3645	0.70	0.69	0.71
Abscond: Total	21046	3983	0.70	0.69	0.71
Commitment: Women	1393	127	0.72	0.67	0.76
Commitment: Men	20059	2907	0.70	0.69	0.71
Commitment: Total	21452	3034	0.70	0.69	0.71

Note. Predicting abscond and new commitment with the General Recidivism Risk Scale.

Table 3.6: AUC results in logistic regression samples (first releases) with 3-year follow-up for felony arrest resulting in a new commitment and 2-year follow-up for the abscond outcome.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Abscond: Women	464	144	0.76	0.72	0.81
Abscond: Men	6968	1523	0.76	0.75	0.77
Abscond: Total	7432	1667	0.76	0.75	0.77
Commitment: Women	647	50	0.73	0.66	0.79
Commitment: Men	6861	1088	0.71	0.69	0.72
Commitment: Total	7508	1138	0.71	0.70	0.73

Note. Predicting abscond and new commitment with the General Recidivism Risk Scale.

Table 3.7: AUC results in logistic regression samples (all releases) with 3-year follow-up for arrest resulting in a new commitment and 2-year follow-up for the abscond outcome.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Abscond: Women	605	237	0.74	0.70	0.78
Abscond: Men	9818	2717	0.73	0.72	0.74
Abscond: Total	10423	2954	0.73	0.72	0.74
Commitment: Women	799	83	0.73	0.68	0.78
Commitment: Men	9733	1937	0.69	0.68	0.70
Commitment: Total	10532	2020	0.69	0.68	0.70

Note. Predicting abscond and new commitment with the General Recidivism Risk Scale.

Table 3.8: AUC results in logistic regression samples (first releases) with 2-year follow-up for arrest outcomes.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Arrest: Women	1352	322	0.73	0.70	0.76
Arrest: Men	17498	4763	0.74	0.73	0.75
Arrest: Total	18850	5085	0.74	0.73	0.75
Felony Arrest: Women	1344	202	0.71	0.67	0.75
Felony Arrest: Men	17369	3427	0.72	0.72	0.73
Felony Arrest: Total	18713	3629	0.72	0.72	0.73
Violent Arrest: Women	1282	40	0.66	0.57	0.75
Violent Arrest: Men	16570	1242	0.72	0.71	0.74
Violent Arrest: Total	17852	1282	0.73	0.71	0.74
VFO Arrest: Women	1279	41	0.67	0.58	0.76
VFO Arrest: Men	16467	1325	0.73	0.72	0.74
VFO Arrest: Total	17746	1366	0.73	0.72	0.75

Note. Predicting arrest and felony arrest with the General Recidivism Risk Scale and violent and VFO arrests with the Violent Recidivism Risk Scale.

Table 3.9: AUC results in logistic regression samples (all releases) with 2-year follow-up for arrest outcomes.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Arrest: Women	1650	448	0.72	0.70	0.75
Arrest: Men	24057	7483	0.73	0.72	0.73
Arrest: Total	25707	7931	0.73	0.72	0.73
Felony Arrest: Women	1638	298	0.71	0.68	0.74
Felony Arrest: Men	23861	5537	0.71	0.70	0.71
Felony Arrest: Total	25499	5835	0.71	0.70	0.71
Violent Arrest: Women	1541	52	0.66	0.59	0.74
Violent Arrest: Men	22378	2004	0.71	0.70	0.72
Violent Arrest: Total	23919	2056	0.71	0.70	0.72
VFO Arrest: Women	1538	53	0.67	0.59	0.74
VFO Arrest: Men	22216	2137	0.72	0.71	0.73
VFO Arrest: Total	23754	2190	0.72	0.71	0.73

Note. Predicting arrest and felony arrest with the General Recidivism Risk Scale and violent and VFO arrests with the Violent Recidivism Risk Scale.

Table 3.10: AUC results in logistic regression samples (first releases) with 3-year follow-up for arrest outcomes.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Arrest: Women	911	286	0.74	0.70	0.77
Arrest: Men	10851	3917	0.74	0.73	0.75
Arrest: Total	11762	4203	0.74	0.73	0.75
Felony Arrest: Women	901	186	0.72	0.68	0.76
Felony Arrest: Men	10733	2858	0.72	0.71	0.73
Felony Arrest: Total	11634	3044	0.72	0.71	0.73
Violent Arrest: Women	855	36	0.63	0.53	0.72
Violent Arrest: Men	10035	1012	0.71	0.69	0.72
Violent Arrest: Total	10890	1048	0.71	0.69	0.73
VFO Arrest: Women	850	39	0.64	0.55	0.73
VFO Arrest: Men	9930	1105	0.72	0.71	0.74
VFO Arrest: Total	10780	1144	0.72	0.71	0.74

Note. Predicting arrest and felony arrest with the General Recidivism Risk Scale and violent and VFO arrests with the Violent Recidivism Risk Scale.

Table 3.11: AUC results in logistic regression samples (all releases) with 3-year follow-up for arrest outcomes.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Arrest: Women	1121	395	0.74	0.71	0.77
Arrest: Men	15124	6161	0.73	0.72	0.73
Arrest: Total	16245	6556	0.73	0.72	0.73
Felony Arrest: Women	1102	272	0.72	0.69	0.76
Felony Arrest: Men	14917	4624	0.71	0.70	0.72
Felony Arrest: Total	16019	4896	0.71	0.70	0.72
Violent Arrest: Women	1016	47	0.64	0.55	0.72
Violent Arrest: Men	13629	1672	0.70	0.69	0.71
Violent Arrest: Total	14645	1719	0.70	0.69	0.71
VFO Arrest: Women	1011	50	0.64	0.57	0.72
VFO Arrest: Men	13464	1824	0.71	0.70	0.72
VFO Arrest: Total	14475	1874	0.71	0.70	0.72

Note. Predicting arrest and felony arrest with the General Recidivism Risk Scale and violent and VFO arrests with the Violent Recidivism Risk Scale.

3.2.2 ROC Methods Summary

We conducted receiver operating characteristic (ROC) analysis to gauge the accuracy of the General Recidivism Risk Score for predicting any arrest, felony offense arrest, absconding, and new commitment offense arrest. We also conducted ROC analysis of the Violent Recidivism Risk Scale predicting violent offense arrest and VFO arrest. The failure was defined as the occurrence of the event of interest within three years of release. The follow-up time was adjusted if the case failed by a competing event such as discharge or return to prison. Time was also adjusted for any periods during which an inmate was held in custody.

Figures 3.2 through 3.4 show the ROC curves for the General Recidivism Risk and Violent Recidivism Risk scales predicting their respective study outcomes for all releases and first releases with at least three years of follow-up time.

The ROC method is useful for guiding the selection of a cutting point on a continuous or ordinal scale to create a binary (no/yes) decision. The ROC curve is a device that can be used to examine the trade-offs between false positives and false negatives. A complete explanation of the ROC approach is provided in the Methods Section. To save space we interpret only the ROC plot for the arrest outcomes.

The top left ROC plot in Figure 3.2 shows ROC curve for the General Recidivism Risk Scale predicting any arrest within three years, for all releases with at least three years of follow-up. The AUC for this model is 0.73. The ROC curve includes a marker that indicates the optimum Recidivism Risk decile score cutting point, if the decile score were used to make a *single* binary (high vs. low) decision. The cutting point that optimizes the sensitivity and specificity is a decile score of D4, i.e., if the decile score is 4 or higher, the case is predicted to fail. Sensitivity (76.9%) is the true positive fraction. Specificity (57.1%) is the true negative fraction, and 1 minus specificity is the false positive fraction (42.9%). The positive predictive value (PV+) is the probability of arrest given a score greater than or equal to D4. The negative predictive value (PV-) is the probability of no arrest given a decile score less than 4. The ROC curve quantifies the accuracy of the risk scale and the predictive value quantifies its clinical value (Pepe, 2003). A useful prediction will have a PV+ that is greater than the base rate and a PV- that is greater than 1 minus the base rate. The base rate for arrest within three years among all releases is 40.4%, so the predictive values (PV+ = 54.8%, PV- = 78.5%) indicate modest clinical value if a single cut were made at D4.

If the default cut at D8 were used (scores at or above 8 are classified as high risk), we get the following results: true positive fraction is 32.2%; false positive fraction is 11.1%; PV+ is 66.1%; and PV- is 65.9%. Compared to the cut at

D5, a cut at D8 yields a lower true positive fraction and slightly lower negative predictive value, but a lower false positive fraction and a higher positive predictive value.

The top right plot in Figure 3.2 shows the ROC curve for the General Recidivism Risk scale deciles score for first releases (parolees and maxouts) with at least three years of follow-up. The AUC for this model is 0.74. The accuracy and predictive value of the General Recidivism Risk scale improve only slightly for predictions in the subset of first releases.

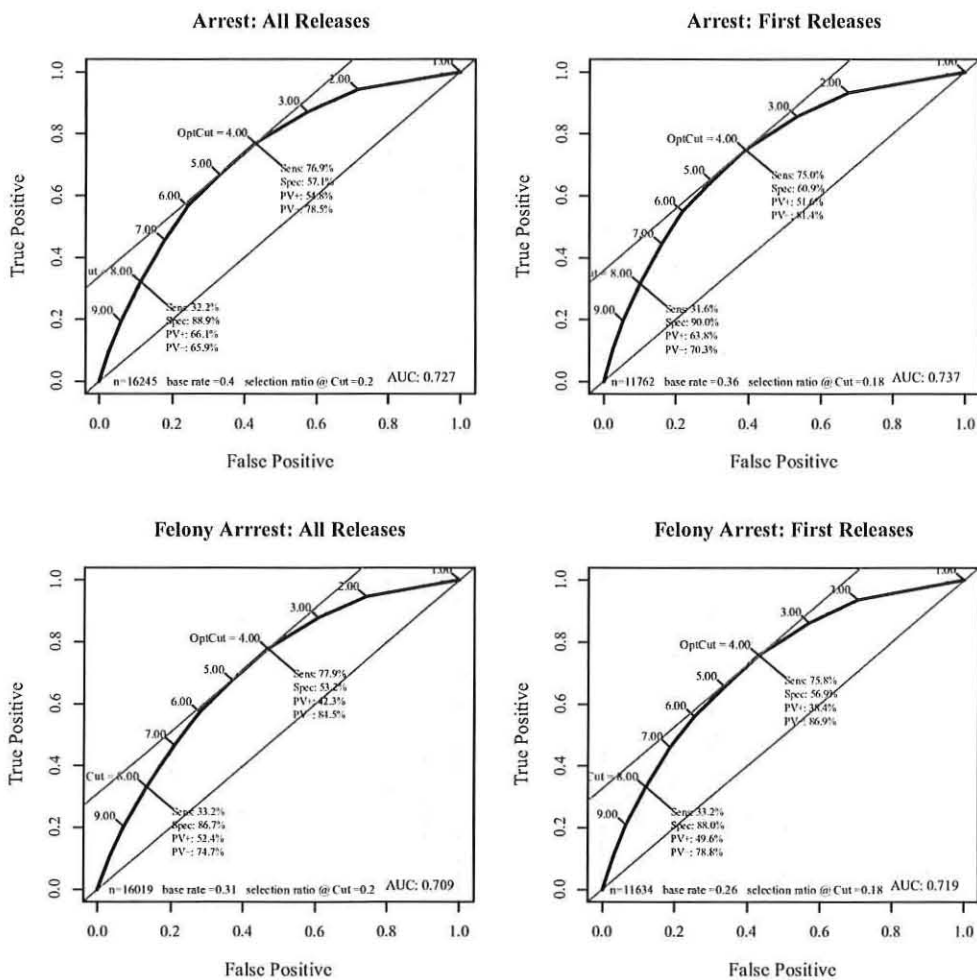


Figure 3.2: Receiver operating characteristic curves for the General Recidivism Risk decile score predicting any arrest and felony arrest within three years for all releases and first releases.

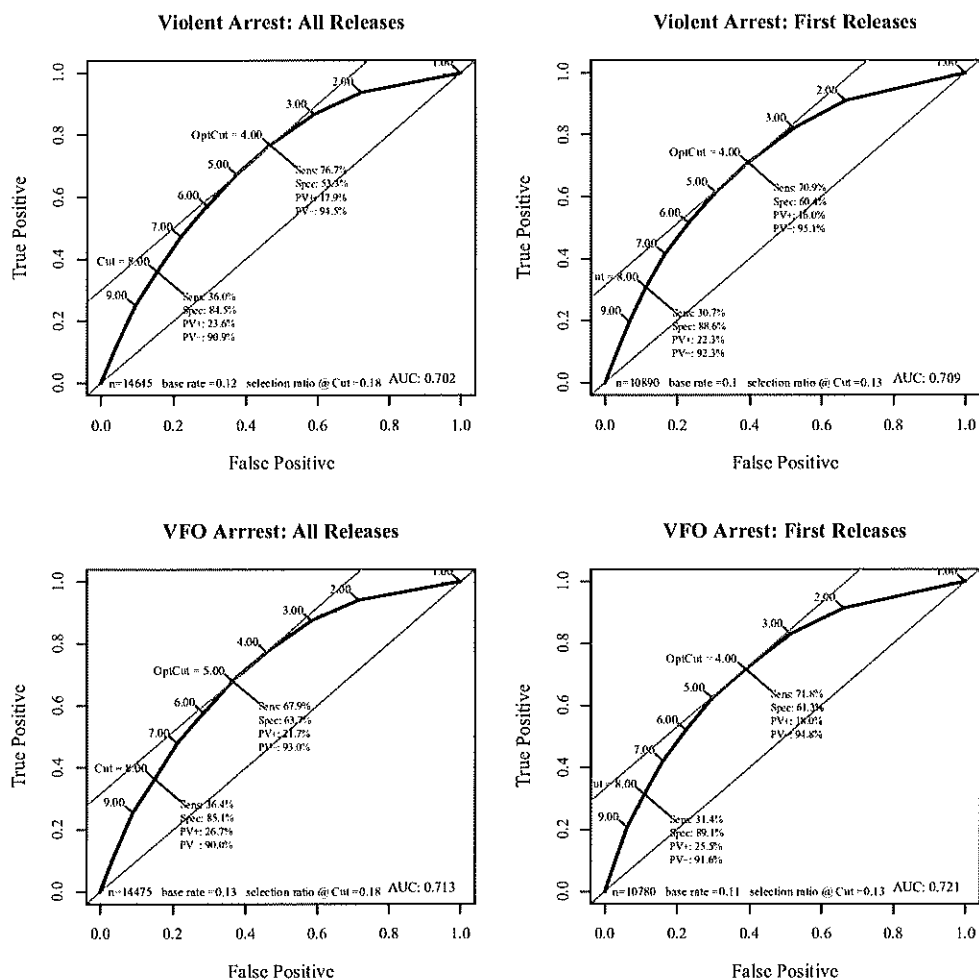


Figure 3.3: Receiver operating characteristic curves for the Violent Recidivism Risk decile score predicting violent offense arrest and violent felony offense (VFO) arrest within three years for all releases and first releases.

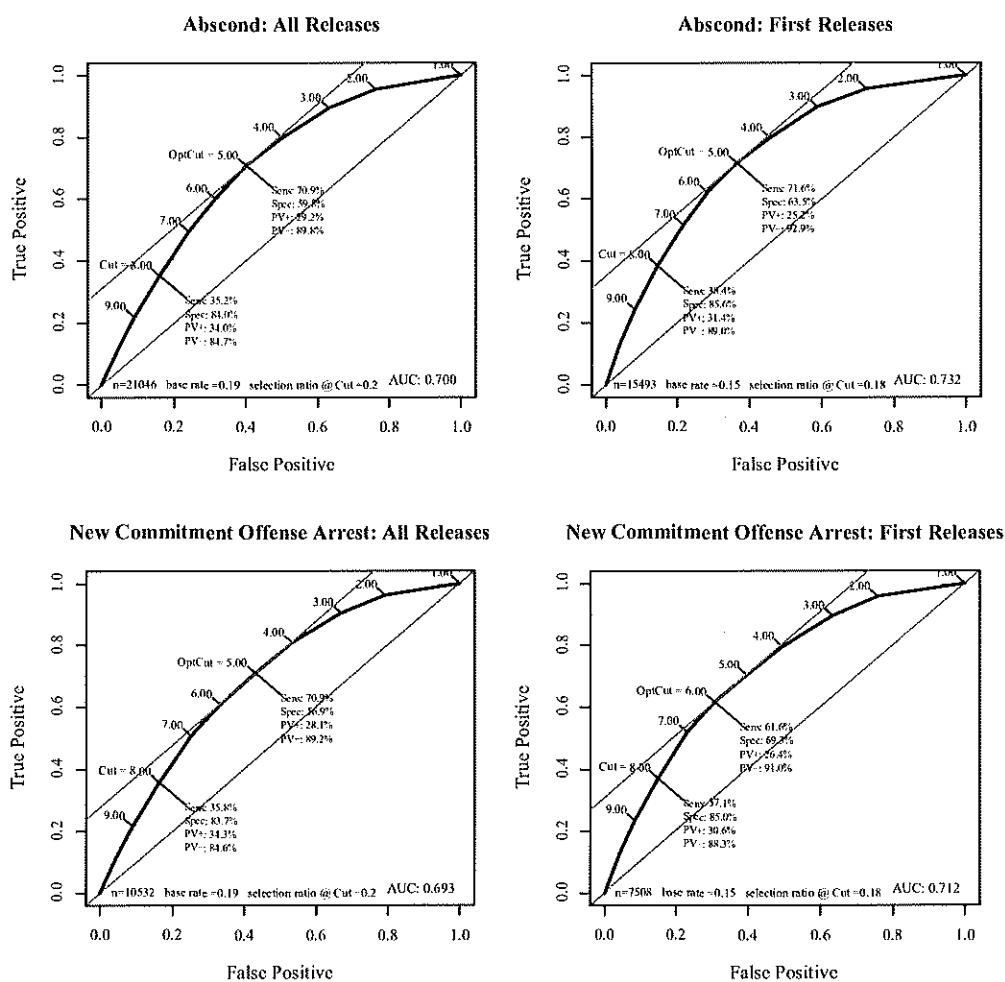


Figure 3.4: Receiver operating characteristic curves for the General Recidivism Risk decile score predicting absconding within 1.5 years and new commitment offense arrest within three years for all releases and first releases.

3.2.3 Failure Probabilities Across the Risk Scale Levels

Figures 3.5 through 3.7 show the percentage failing in the Recidivism Risk Scale levels and Violent Recidivism Risk Scale levels. The top plots in Figure 3.5 show the observed probabilities of a new arrest within three years of release for all releases (top left) and first releases (top right). Only those cases that had at least three years of potential follow-up time are included. For all releases the observed proportion with an arrest within three years is 0.25 in the low level, 0.52 in the medium level, and 0.66 in the high level. The base rate for arrest within three years of release is 0.4 for all releases ($n = 16,245$) and 0.36 for first releases ($n = 11,762$).

The plots are annotated with the odds ratio of the high versus low levels of the risk scales. Overall, the odds ratios and failure probabilities indicate that the levels of the risk scales effectively divide the reentry outcomes sample into groups with distinct base rates of failure.

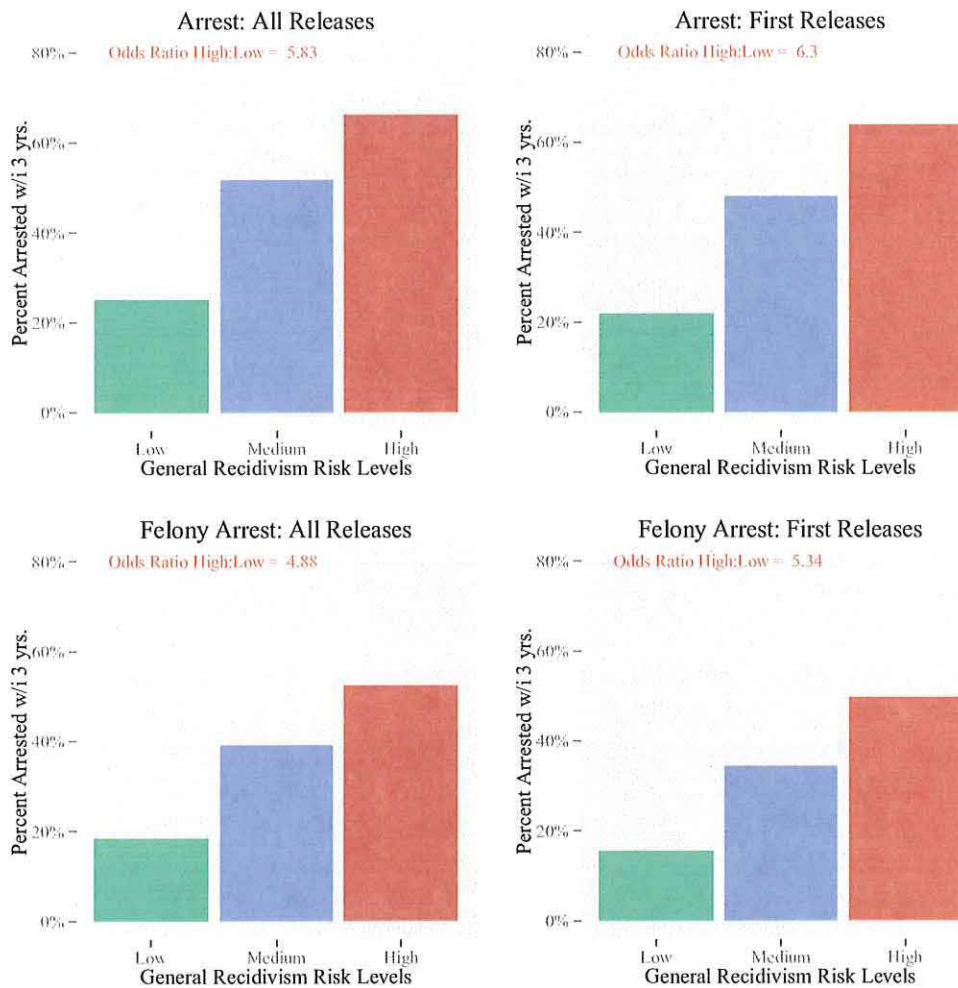


Figure 3.5: Percent of inmates arrested for any offense and percent arrested for a felony offense within three years across the General Recidivism Risk Levels for all releases and first releases.

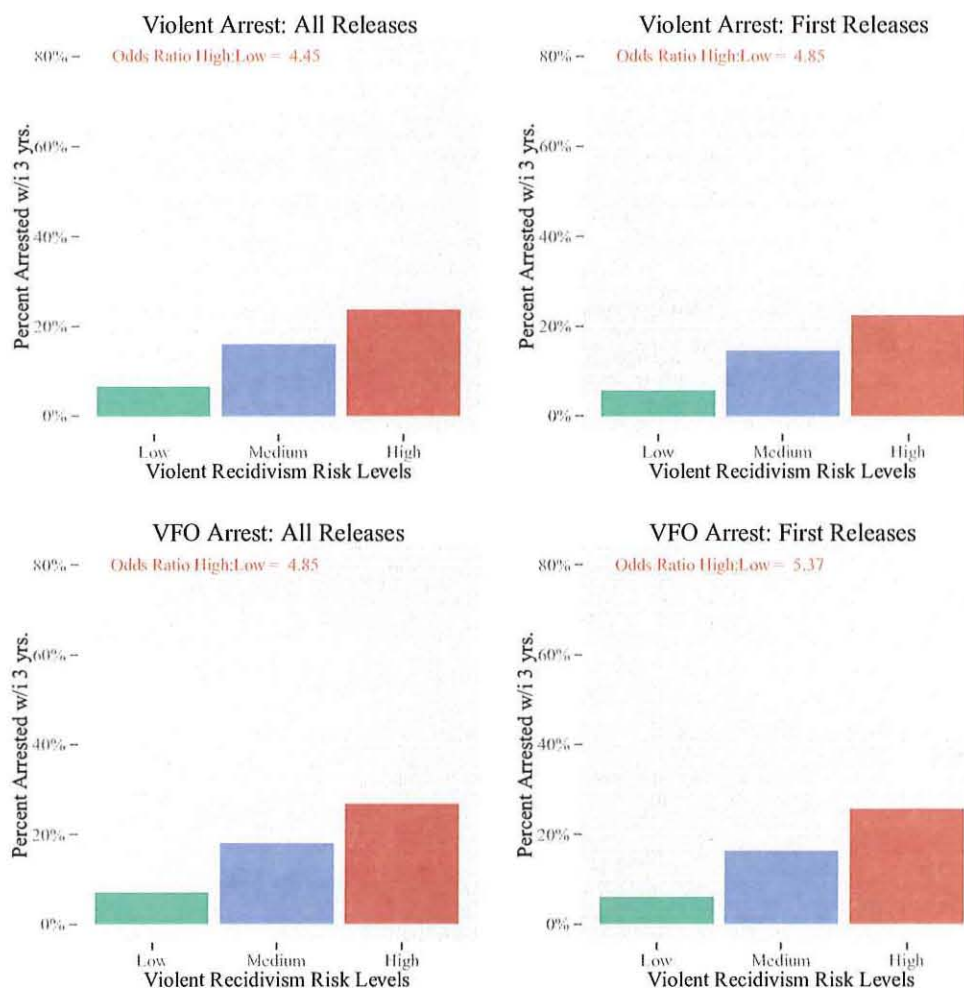


Figure 3.6: Percent of inmates arrested for a violent offense and percent arrested for a violent felony offense (VFO) within three years across the Violent Recidivism Risk Levels for all releases and first releases.

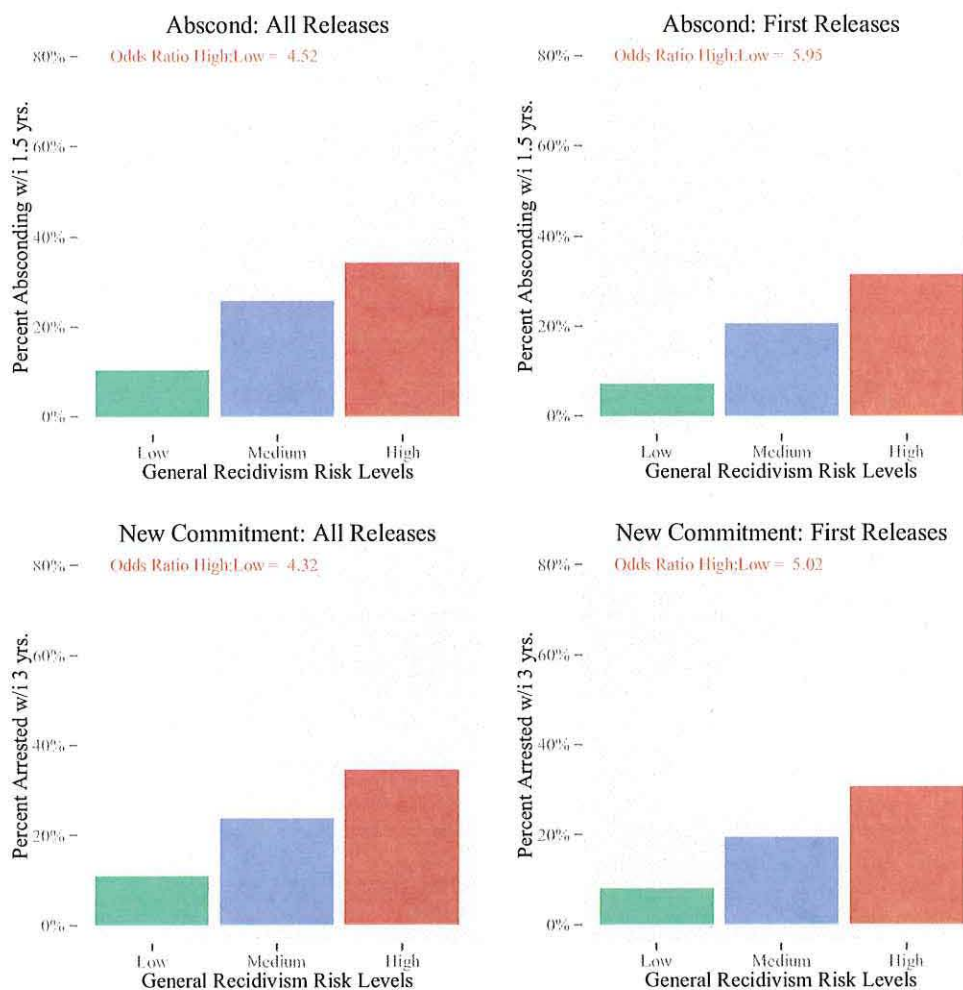


Figure 3.7: Percent of inmates absconding within 1.5 years across the General Recidivism Risk Levels and percent of inmates arrested for a new commitment offense within three years across the General Recidivism Risk Levels for all releases and first releases.

3.2.4 Survival Plots

Figures 3.8 through 3.10 show plots of the naive Kaplan-Meier survival curves and cumulative incidence functions within the levels of the risk scales for all releases. In typical survival data without competing events the Kaplan-Meier statistic is used as an estimate of survival or failure probability ($1 - \text{KM}$) at different time points. However, in the context of competing risks, $1 - \text{KM}$ is not a proper or interpretable failure probability. Usually $1 - \text{KM}$ will be biased upward. When modeling competing risks, the recommended approach to displaying results is to plot the cumulative incidence functions. If failure from competing events is low, then the Kaplan-Meier statistic will be similar to the cumulative incidence. The higher the competition from other events the higher the upward bias of the $1 - \text{KM}$ failure probability.

The effect of competing events on the probability of recidivism is relevant for the conduct of the decision analytics study being conducted for the MDOC. The objective of the decision analytics study is to develop decision thresholds for the risk scales based on predicted recidivism probabilities. Thus it is important for the conduct of that study to understand how competing events alter the probability of observing recidivism events of interest.

A comparison of the Kaplan-Meier curves and cumulative incidence curves for absconding shown in the top row of Figure 3.10 provides a good example of the effect of competing failures on the probability estimate. Events that compete with absconding were shown previously in Figure 3.1. Censoring, due mostly to parole discharge, is also reflected in the size of the risk set over time shown at the top of the Kaplan-Meier plot in Figure 3.10. The size of the risk set drops off sharply around 540 days, coinciding with the timing of parole discharge. Thus, the probability of absconding is altered substantially by failures from other competing events, especially parole discharge, and consequently, as shown in 3.10, the $1 - \text{KM}$ estimate is much higher than the cumulative incidence.

The plots show good separation in fitted probabilities between the high, medium, and low levels of the General Recidivism Risk Scale using the default cut points at D5 and D8 predicting new commitment offense arrest (Figure 3.10). The plots also indicate good separation between the high, medium, and low levels of the Violent Recidivism Risk Scale predicting VFO arrest (Figure 3.9). However, for most of the outcomes and risk scales, using the default cuts at D5 and D8, the separation between the medium and high levels is less than the separation between the low and medium levels.

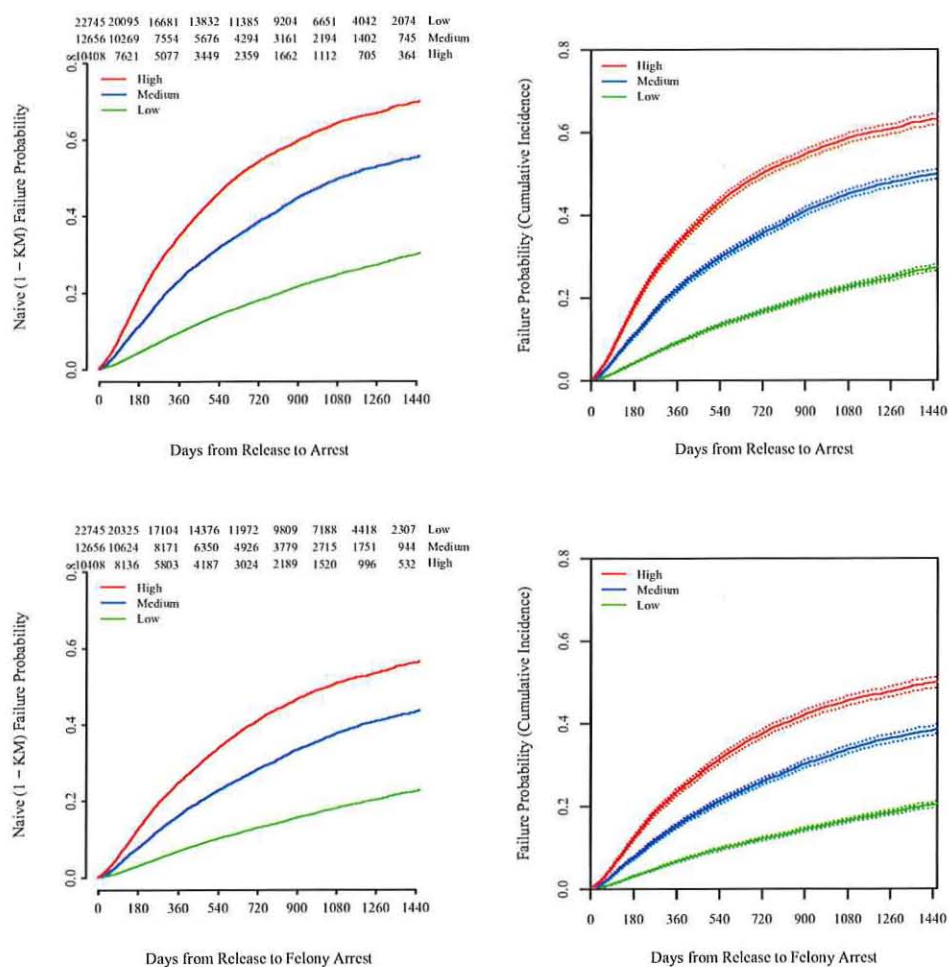


Figure 3.8: Kaplan-Meier failure probability (left panels) and cumulative incidence (right panels) of arrest and felony arrest for all releases within the levels of the General Recidivism Risk scale with cuts at D5 and D8.

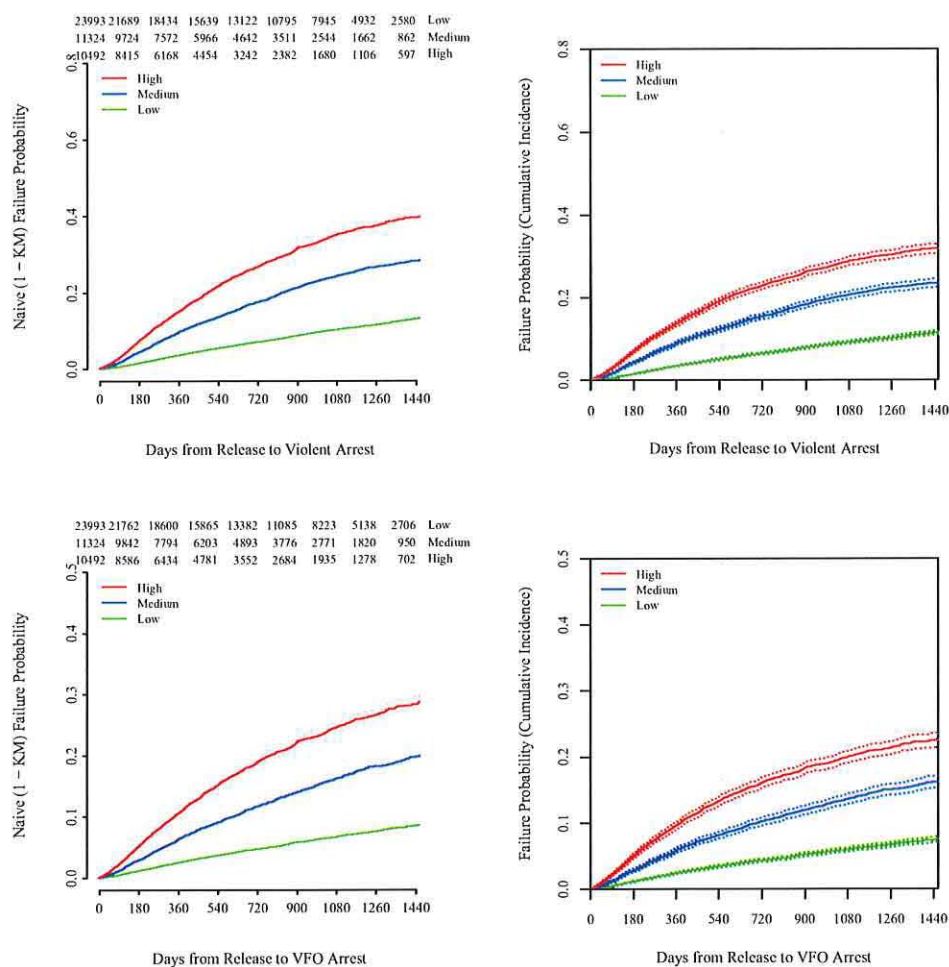


Figure 3.9: Kaplan-Meier failure probability (left panels) and cumulative incidence (right panels) of violent arrest and violent felony offense (VFO) arrest for all releases within the levels of the Violent Recidivism Risk Scale with cuts at D5 and D8.

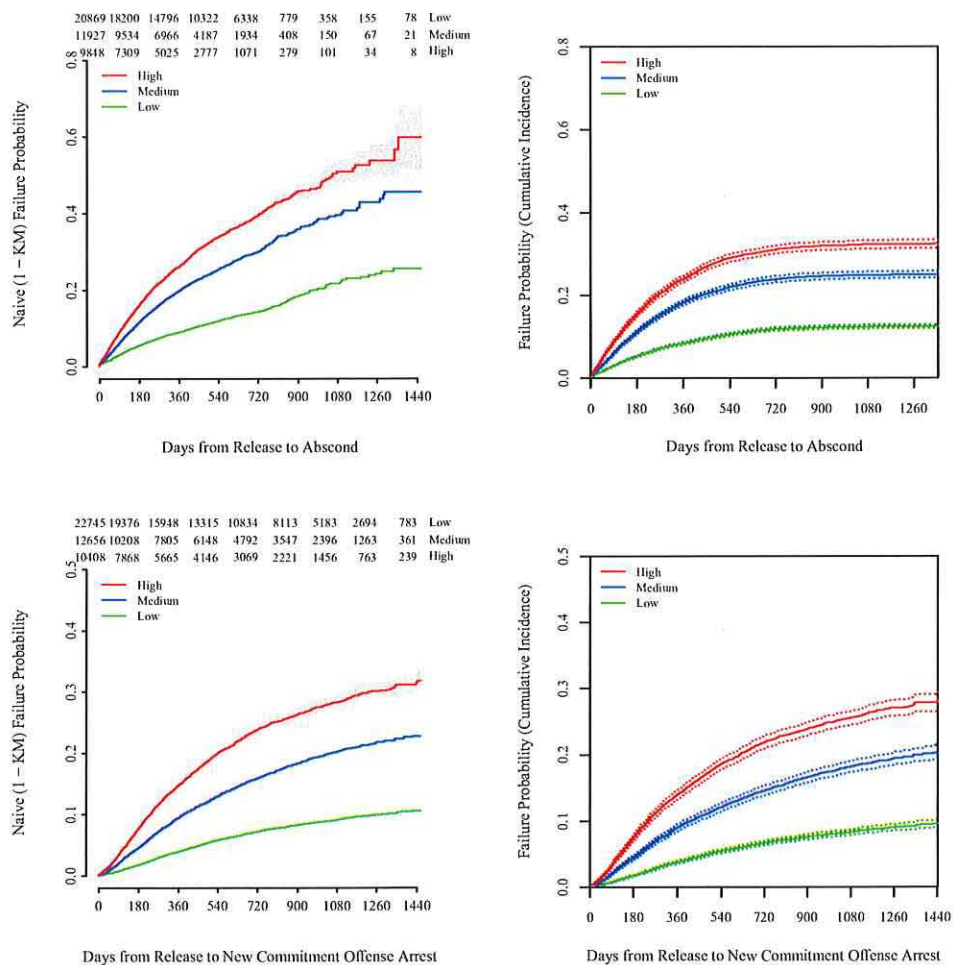


Figure 3.10: Kaplan-Meier failure probability (left panels) and cumulative incidence (right panels) of absconding and new commitment offense arrest for all releases within the levels of the General Recidivism Risk Scale with cuts at D5 and D8.

Appendix A

2012 Reentry Study AUC Results

Table A.1: AUC results in logistic regression samples (first releases) with 2-year follow-up for arrest-resulting-in-conviction outcomes and 1.5-year follow-up for the abscond outcome.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Abscond: Women	755	150	0.73	0.68	0.77
Abscond: Men	9539	1350	0.73	0.72	0.74
Abscond: Total	10294	1500	0.73	0.72	0.74
Felony: Women	660	93	0.73	0.68	0.78
Felony: Men	6896	1448	0.70	0.69	0.72
Felony: Total	7556	1541	0.70	0.69	0.72
VFO: Women	613	17	0.62	0.47	0.77
VFO: Men	6268	452	0.70	0.68	0.72
VFO: Total	6881	469	0.70	0.68	0.73
Commitment: Women	639	40	0.72	0.64	0.79
Commitment: Men	6746	914	0.71	0.69	0.73
Commitment: Total	7385	954	0.71	0.70	0.73

Note. Predicting abscond, felony, and new commitment with the General Recidivism Risk Scale and VFO with the Violent Recidivism Risk Scale.

Table A.2: AUC results in logistic regression samples (all releases) with 2-year follow-up for arrest-resulting-in-conviction outcomes and 1.5-year follow-up for the abscond outcome.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Abscond: Women	941	239	0.71	0.67	0.74
Abscond: Men	13312	2487	0.70	0.69	0.71
Abscond: Total	14253	2726	0.70	0.69	0.71
Felony: Women	834	155	0.73	0.69	0.77
Felony: Men	9916	2514	0.68	0.67	0.69
Felony: Total	10750	2669	0.69	0.68	0.70
VFO: Women	751	21	0.65	0.51	0.78
VFO: Men	8737	773	0.70	0.68	0.71
VFO: Total	9488	794	0.70	0.68	0.72
Commitment: Women	799	69	0.73	0.67	0.79
Commitment: Men	9644	1632	0.69	0.67	0.70
Commitment: Total	10443	1701	0.69	0.68	0.70

Note. Predicting abscond, felony, and new commitment with the General Recidivism Risk Scale and VFO with the Violent Recidivism Risk Scale.

Table A.3: AUC results in logistic regression samples (first releases) with 2-year follow-up for arrest outcomes.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Arrest: Women	968	251	0.73	0.70	0.77
Arrest: Men	11721	3504	0.73	0.72	0.74
Arrest: Total	12689	3755	0.73	0.72	0.74
Felony Arrest: Women	963	169	0.72	0.68	0.76
Felony Arrest: Men	11672	2699	0.72	0.71	0.73
Felony Arrest: Total	12635	2868	0.72	0.71	0.73
Violent Arrest: Women	910	108	0.65	0.59	0.70
Violent Arrest: Men	10956	1855	0.70	0.69	0.71
Violent Arrest: Total	11866	1963	0.70	0.69	0.71
VFO Arrest: Women	906	34	0.65	0.55	0.75
VFO Arrest: Men	10888	1080	0.72	0.70	0.73
VFO Arrest: Total	11794	1114	0.72	0.71	0.74

Note. Predicting arrest and felony arrest with the General Recidivism Risk Scale and violent and VFO arrests with the Violent Recidivism Risk Scale.

Table A.4: AUC results in logistic regression samples (all releases) with 2-year follow-up for arrest outcomes.

outcome: sample	n	events	AUC	Lower 95% CI	Upper 95% CI
Arrest: Women	1199	357	0.73	0.70	0.76
Arrest: Men	16426	5635	0.72	0.71	0.73
Arrest: Total	17625	5992	0.72	0.71	0.73
Felony Arrest: Women	1191	256	0.72	0.69	0.76
Felony Arrest: Men	16327	4431	0.70	0.70	0.71
Felony Arrest: Total	17518	4687	0.71	0.70	0.71
Violent Arrest: Women	1103	155	0.66	0.62	0.71
Violent Arrest: Men	15016	2915	0.69	0.68	0.70
Violent Arrest: Total	16119	3070	0.69	0.68	0.70
VFO Arrest: Women	1097	45	0.67	0.59	0.76
VFO Arrest: Men	14893	1775	0.71	0.70	0.72
VFO Arrest: Total	15990	1820	0.71	0.70	0.72

Note. Predicting arrest and felony arrest with the General Recidivism Risk Scale and violent and VFO arrests with the Violent Recidivism Risk Scale.

Bibliography

- Andrews, D. A., & Bonta, J. (1995). *LSI-R: The Level of Service Inventory-Revised*. Toronto, Canada: Multi-Health Systems.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or needs assessment. *Crime and Delinquency*, 52(1), 7-27.
- Barbaree, H. E., Seto, M., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, 28(4), 490-521.
- Barnoski, R. (2006). *Sex offender sentencing in Washington State: Predicting recidivism based on the LSI-R*. (Tech. Rep.). Olympia, WA: Washington State Institute for Public Policy. Retrieved from <http://www.wsipp.wa.gov/rtpfiles/06-02-1201.pdf>
- Barnoski, R., & Aos, S. (2003). *Washington's offender accountability act: An analysis of the Department of Corrections' risk assessment (document no. 03-12-1202)* (Tech. Rep.). Olympia, WA: Washington State Institute for Public Policy.
- Brennan, T., & Dieterich, W. (2008). *Michigan Department of Corrections Core COMPAS pilot study: One-year follow-up* (Tech. Rep.). Traverse City, MI: Northpointe Institute for Public Management.
- Brennan, T., Dieterich, W., & Khan, M. N. (2009). *Testing the predictive validity of the Core COMPAS risk scales: Phase II results from an outcomes study conducted for the Michigan Department of Corrections* (Tech. Rep.). Traverse City, MI: Northpointe Institute for Public Management.
- Dahle, K. P. (2006). Strengths and limitations of actuarial prediction of criminal re-offence in a German prison sample: A comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law and Psychiatry*, 29(5), 431-442.
- Dieterich, W., Brennan, T., & Breitenbach, M. (2010). *Testing the predictive validity of the COMPAS Reentry risk scales: Results from an outcome study conducted for the Michigan Department of Corrections*.

- (Tech. Rep.). Traverse City, MI: Northpointe Institute for Public Management.
- Dieterich, W., Oliver, W., & Brennan, T. (2011). *Predictive validity of the Reentry COMPAS Risk scales: An outcomes study with extended follow-up conducted for the Michigan Department of Corrections*. (Tech. Rep.). Northpointe Inc.
- Dieterich, W., Oliver, W. L., & Brennan, T. (2012). *Predictive validity of the COMPAS Reentry risk scales: An outcomes study conducted for the Michigan Department of Corrections: Updated results on the 2008-2010 release sample*. (Tech. Rep.). Northpointe Inc.
- Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446), 496-509.
- Flores, A. W., Lowenkamp, C. T., Smith, P., & Latessa, E. J. (2006). Validating the Level of Service Inventory-Revised on a sample of federal probationers. *Federal Probation*, 70(2), 44-78.
- Grann, M., Belfrage, H., & Tengstrom, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior*, 27, 97-114.
- Hare, R. D. (1991). *Manual for the Hare Psychopathy Checklist-Revised*. Toronto, Canada: Multi-Health Systems.
- Kroner, C., Stadtland, C., Eidt, M., & Nedopil, N. (2007). The validity of the violence risk appraisal guide (VRAG) in predicting criminal recidivism. *Criminal Behaviour and Mental Health*, 17(2), 89 - 100.
- Northpointe Inc. (2013). *MDOC VFO and Non-VFO Reentry Risk Scores: Scale sets and scoring rules*. (Tech. Rep.). Traverse City, MI: Northpointe Inc.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Oliver, W. L., Dieterich, W., & Brennan, T. (2013a). *New COMPAS Reentry Risk Scales for predicting VFO and Non-VFO arrests* (Tech. Rep.). Traverse City, MI: Northpointe Inc.
- Oliver, W. L., Dieterich, W., & Brennan, T. (2013b). *Setting decision cut points for the new VFO and Non-VFO risk scales: A research report on the use of predicted probabilities in the MDOC reentry application*. (Tech. Rep.). Traverse City, MI: Northpointe, Inc.
- Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. New York: Oxford University Press.
- Pintilie, M. (2007). Analysing and interpreting competing risk data. *Statistics in Medicine*, 26, 1360 - 1367.
- Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, 26,

2389 - 2430.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.