CENTER FOR APPLIED RESEARCH & ANALYSIS

New Mexico Corrections Department (NMCD) External Classification Validation Study

Date: March 2024

Alex Severson, Ph.D.
Paul Guerin, Ph.D.
Reanna Sanchez, M.A.
Cris Moore, Ph.D.

Prepared for:
New Mexico Corrections Department

Table of Contents

Introduction

Having an effective prison classification process is essential to optimize institutional safety and to ensure prisons are responsive to the needs of inmates and changes in demographics, policy, and resources. Because maintaining institutional safety is a primary concern of correctional institutions (Berk, Kriegler, and Baek 2006; Steiner and Kane 2016), inmate misconduct is costly (Lovell and Jemelka 1996; Tewksbury et al., 2014; Morgan 2018), and misconduct in prison predicts recidivism (Cochran et al., 2014; Cochran and Mears, 2017), it is important to evaluate whether existing classification tools effectively segregate prison populations by misconduct risk (Labrecque et al., 2020; Long 2020). Validation studies that evaluate predictive performance can provide this understanding (Gottfredson and Moriarty 2006; Singh et al., 2013; Desmarais et al., 2018).

The initial custody classification tool administered at intake and the reclassification tool used by the New Mexico Corrections Department (NMCD) to evaluate institutional misconduct have not been validated since their introduction in 2003. Accordingly, the primary aims of ISR's Phase III report are to (1) assess how classification officers and supervisory staff perceive the classification process and (2) validate the existing tools using quantitative methods to evaluate the extent to which the tools predict misconduct.

In what follows, we discuss the context and development of the NMCD's classification policy, briefly review the peer-reviewed literature on institutional misconduct and risk assessment validation, describe the results of observations of the classification process and five focus groups designed to assess how classification staff at the NMCD perceived the classification process, present the results of a validation analysis designed to see how well the classification tools predicted misconduct, and provide recommendations to the NMCD to improve predictive accuracy and adoption of the tools. The present report on the NMCD classification system described in NMCD policy CD-080100 extends work ISR completed in Phase I and Phase II studies.

Background on Prison Classification

Classification is an important process for separating offender populations based on predictive elements of risk and needs. The goal of classification is to ensure that inmates are assigned to a facility that corresponds to their custody level and to separate inmates based on the threats they pose to one another, correctional staff, and the public. Classification processes typically consider factors, such as material, psychological, and relational insufficiencies, that may underlie these risks (Austin & Hardyman, 2004).

The NMCD uses both an external and internal classification system. External classification refers to the classification of offenders entering or remaining within the prison system and involves assigning inmates to a facility that corresponds to their custody level. Internal classification refers to the system used to determine which cell, housing unit, and programming opportunities (e.g., education, vocational, counseling, and work assignments) inmates are assigned to following external classification. The external classification system used by the NMCD determines facility placements. In New Mexico, prisons are ranked based on a four-level system. Level I is the least restrictive level while Level IV is the most restrictive. ISR's Phase III research pertains to the external classification system's initial and reclassification processes and scoring forms.

Within the NMCD, the external classification process consists of three parts: (1) the reception and diagnostic process for both men and women which occurs at Central New Mexico Correctional Facility for men and Western New Mexico Correctional Facility for women; (2) the initial external classification process in which a classification officer scores an inmate based on their case and make a recommendation on custody level; and (3) the reclassification process that takes place at each facility.

In 2003, the NMCD implemented the current external classification system for New Mexico inmates. The tool was developed using professional judgment and expert consultation to select cutoff scores for different factors and custody levels. The initial classification tool used by the NMCD contains eight factors: history of institutional violence, severity of current conviction, escape history, prior felony convictions, severity of prior convictions, alcohol and drug abuse, age at assessment, and gang membership or activities in the prior ten years. The reclassification tool used by the NMCD contains nine factors: history of institutional adjustment and violence, severity of current conviction, escape history, prior felony convictions, disciplinary convictions received, history of disciplinary misconduct, program and work performance, current age at assessment, and gang membership and activities in the past ten years (see Appendix A for the text and scoring of both the initial and reclassification scoring forms).

The initial and reclassification tools used by the NMCD were not modified following completion of Phase I and II studies which, among other things, recommended the creation of gender-responsive tools and the modification of factor scoring (e.g., the creation of separate classification forms for male and female inmates and the inclusion of different variables and custody level scoring for female inmates). Additionally, Phase I and II reports recommended the inclusion of history of mental illness as a scored factor within the classification tools instead of a factor which automatically triggered a mandatory override into a higher custody level due to limited mental health services available at lower-custody level facilities.

Each response option for each factor on NMCD's initial classification and reclassification tools has a numeric score, and an inmate's overall score on the tool is the sum of their scores across all factors. Higher scores indicate a greater risk of misconduct. For example, the severity of current conviction factor on both the initial classification and reclassification tool is a factor where an inmate can receive between one to seven points based on the severity level of their current conviction (e.g., 1 = Lowest Severity; 3 = Moderate Severity; 5 = High Severity; 7 = Highest Severity). Scores for each item on the initial and reclassification tool are summed to generate a classification score which can range from a -1 to +32 for the initial classification tool and from -7 to 21 for the reclassification tool. Based on these scores and on custody level breaks within them which vary by sex, inmates receive a score-based custody level which, in most cases, determines facility placement and assigned custody levels from Level I facilities (Least restrictive custody; lowest presumptive misconduct risk) to Level IV facilities (Most restrictive custody, highest presumptive misconduct risk).

An inmate's final classification level and facility placement may not reflect the score they receive on the initial or reclassification tool for two reasons. First, if inmates exceed a specific score threshold on the first three tool factors on either the initial or reclassification tool (i.e., an inmate scores more than 9 points cumulatively on the history of institutional adjustment/violence, severity of current conviction, and escape history factors), as a matter of NMCD policy, these inmates are automatically classified as Level IV given

perceived flight-risk and disorder concerns associated with these three factors. Second, differences between an inmate's score and final facility placement may occur due to the NMCD's use of mandatory and discretionary overrides. Mandatory overrides, which override inmates from lower custody levels into higher custody levels, are automatically triggered through NMCD policy for a variety of reasons including the type of crime an inmate has committed or specific institutional considerations (e.g., if an inmate has been convicted of a crime involving assaultive sexual conduct or has been convicted of a crime involving violence to a child, they are automatically overridden into a higher custody level; similarly, if an inmate meets certain acuity levels for mental illness, they are often overridden into higher custody levels to secure access to appropriate mental health services, medication, and programming opportunities). In the admission cohort we analyzed, 93% of all overrides were mandatory overrides with a plurality increasing the classification level due to mental health and medical restrictions.

Discretionary overrides occur on a case-by-case basis at the discretion of correctional officers and are used in cases where there are known behavioral or management problems and known or suspected gang affiliation, among other things. Whereas mandatory overrides exclusively result in placement of inmates into more restrictive custody levels, discretionary overrides can either result in placement above or below their scored level, conditional on the override reason. As a matter of policy, mandatory overrides must be approved by supervisors whereas discretionary overrides must be approved by the Central Classification Bureau. We provide a detailed description of the charges and events which trigger mandatory and discretionary overrides in Appendix B.

We comment more on the frequency of the NMCD's use of mandatory and discretionary overrides (e.g., overrides occur in approximately 45% of all classification records within the admission cohort) and the potential challenges their frequent issuance use present to the present validation efforts in the *Recommendations* section of the present report. We also invite readers to examine the linked 2020 Legislative Finance Committee report on NMCD's inmate classification system which details a similar set of concerns with the NMCD's use of mandatory overrides.

Literature Review on Prison Misconduct

Prison misconduct consists of any rule-violating behavior within inmates (Steiner and Wooldredge 2014; Butler et al., 2021a). Studies on the prevalence of prison misconduct suggest that much like crime in the general population, a small subset of inmates is responsible for most serious prison misconduct, though nontrivial proportions of inmates typically engage in institutional misconduct throughout their confinement period (Steiner and Cain 2016; Butler et al., 2021a; Butler et al. 2021b). For national context on rates of prison misconduct, data taken from the most recently available nationally-representative correctional survey, the 2014 Annual Survey of Jails, suggests that correctional facilities experience high rates of misconduct, averaging approximately 949 incidents of misconduct per 1,000 inmates (BJS 2015). A study by Sorensen and Cunningham (2007) estimated that approximately 25% of inmates engaged in some form of rule violation with violence potential during their incarceration period.

Identifying inmates who are at the highest risk of engaging in institutional misconduct, particularly serious violent misconduct, can help correctional authorities make informed decisions about where to house these inmates and allocate resources to reduce the risk of violence (e.g., introducing educational or

substance use interventions conditional on inmate needs within each custody level) (Andrews, Bonta, and Hoge 1990). This can improve safety for prison staff, visitors, and other inmates. However, it can be difficult to determine if the custody level an inmate is placed in is the cause of reduced misconduct rates, as higher-custody level facilities may reduce misconduct risk by limiting an inmate's mobility or interactions with others yet higher risk inmates may engage in more misconduct (Harer and Langan 2001; Worrall and Morris 2011). Previous studies report mixed results on the relationship between custody level and misconduct rates. To evaluate this question, we used a quasi-experimental regression discontinuity (RD) design (Berk et al. 2003; Camp and Gaes 2005; though see Worrall and Morris 2011),

Criminologists have studied the individual and institutional factors that predict institutional misconduct. Although there is no meta-analytic work on this topic, previous research has identified common factors that predict misconduct. A 2014 systematic review of 98 studies which investigated the correlates of institutional misconduct reported that institutional misconduct can be predicted from age, prior criminal record, prior misconduct, and prison characteristics such as security levels (Steiner et al., 2014). Within the subset of factors (e.g., typically factors such as age, race, sex, marital status, gang membership, prior arrests, violent offending, and past drug abuse or drug-related problems), age and sex appear to be most consistently predictive of misconduct risk across studies, with younger inmates, specifically those within the 18-25-year age cohort and typically males, having the highest likelihood of engaging in serious violent misconduct (Valentine, Mears, and Bales 2015; Blowers and Blevins 2015). Prior criminal history, which includes measurements such as the seriousness of offense and the recency of offense, recent prison misconduct, race-ethnicity, sentence length, specific forms of gang affiliation, and cognitive ability are also well-supported individual-level predictors of institutional misconduct (Rocheleau, A. M. 2014; Drury and DeLisi, 2010; Diamond et al., 2012; Pyrooz et al., 2016).

It is important to keep in mind that certain predictors may be more or less reliable in identifying the risk of different types of misconduct. For example, according to research by Walters and Crawford (2013), factors such as age, marital status, street gang affiliation, criminal thinking, prior drug abuse, and criminal history may be more predictive of high or high-moderate severity misconduct but may not be as useful for predicting less severe misconduct. Similarly, Walters and Crawford (2014) suggest that the risk of general and aggressive institutional misconduct is highest when an inmate has a history of both mental illness and criminal violence, indicating that the impact of mental illness on violent misconduct is influenced by an inmate's prior violence history.

Other factors that may impact an inmate's likelihood of engaging in misconduct include substance abuse history, mental and physical health issues, and participation in prison programming (Bonner et al., 2017; Semezna and Grosholz 2019). However, the relationship between these factors and misconduct is not always clear due to previous studies' sample sizes, mixed results, and limitations in how outcomes are measured (e.g., not all forms of mental illness equally contribute to misconduct and similarly, not all forms of prison programming contribute uniformly to reductions in misconduct rates) (Severson et al., 2019; Papalia et al., 2019). Institutional-level factors, such as facility security and prison culture, have also been linked to misconduct risk, but incorporating these factors into classification tools can be difficult due to the need to standardize tools across sites, concerns about inmate civil liberties, and implementation challenges (List 2023; Tahamont 2019; Bosma et al., 2020; Kigerl and Hamilton 2016). Some previous research suggests that there may be differences in the frequency, type, and predictors of

institutional misconduct between male and female inmates. For instance, some studies suggest that female inmates engage in less misconduct than male inmates, and the predictors of female misconduct are different from those of male misconduct. Celinska and Sung (2017) present evidence that: "…non-Hispanic ethnicity, a high school diploma, prior physical abuse, drug offenses, major psychiatric disorders, participation in substance use treatment and in religious activities affect only male inmates, while the number of visits in past month affects only female inmates" (p. 237). Recent studies suggest that psychopathy and impulsivity are uniquely important predictors of violent misconduct among female inmates (Thomson et al. 2016; Thomson et al., 2019). Additionally, female inmates may require targeted cognitive-behavioral and relational interventions to reduce misconduct risk which may be related to interpersonal social interactions and network features (Matthews and Hubbard 2008; Wright et al., 2012). For these reasons, some advocate for the creation of sex-specific risk assessment tools that consider sex differences in misconduct type, form, and predictors (Ho and Rocheleau 2020).

However, other research implies that arguments for developing distinct risk assessments for male and female inmates may be overstated. For instance, in a 2014 article, "Sex Differences in the Predictors of Prisoner Misconduct", Steiner and Wooldredge (2014) present evidence that, "…there are far more similarities than differences in the predictors of misconduct among men versus women" (433). A 2014 systematic review on the causes and correlates of inmate misconduct reported that, of 20 studies on prison misconduct published between 1980 and 2013 which included sex as a variable in models predicting misconduct, 50% of studies included found that the effect of sex on misconduct was not statistically significant and 23% of studies found inverse effects (i.e., that female inmates engaged in more misconduct than male inmates) (Steiner, Butler, and Ellison 2014). A 2023 article in *The Prison Journal*, "Serious, Violent, and Chronic Prison Misconduct: Are the Predictors the Same for Women and Men?" suggests that in terms of sex differences in the predictors of serious, violent, and chronic prison misconduct, there were more similarities than differences between male and female inmates, though it is worth noting that the authors identified some differences in the number of predictors of misconduct at the 99th percentile distribution of misconduct (Logan et al., 2023). Given the mixed literature on the relationship between sex and misconduct, it is worth evaluating the unique misconduct profiles of male and female inmates within the NMCD system to determine whether there are meaningful differences in misconduct rates, types, and predictors and, in consequence, whether there is a need to create sex-responsive tools. Classification tool creators need to consider these factors to develop individualized, predictive assessments and should be aware that the results from prior validation studies may not be applicable to all sites or times (Lovins et al., 2018; Wang et al., 2022).

Risk Assessment Research

Research shows that actuarial[1] risk assessments are more accurate than risk assessments formed using professional judgment in classifying inmates based on their risk of misconduct (Hanson and Morton-Bourgon 2009; Ting et al., 2018; Lin et al., 2020). To create effective actuarial risk assessment tools, correctional authorities and scale-developers should include factors in classification tools that are predictive of misconduct while excluding factors that are non-predictive or down-weighting scores for

[1] In the context of risk assessments, actuarial refers to a method of assessing risk that is based on statistical models and algorithms. Actuarial risk assessments use data from past cases to identify factors that are associated with a higher or lower risk of a particular outcome, such as reoffending or misconduct. These factors are then used to assign scores to individuals, which are used to determine their risk level. The use of actuarial risk assessments is intended to improve the accuracy and consistency of risk assessments, by removing subjective judgments and relying instead on objective data and statistical models.

weakly predictive factors. Failure to do so can introduce statistical biases, such as multicollinearity and model overfitting, which can undermine the tool's predictive performance. As Labrecque (2022) notes:

> *Additionally, a growing body of studies discusses the importance of validating and norming risk assessments on local populations (Lovins et al., 2018), further suggesting that weighting the scores of the individual risk items by their predictive strength is one way to maximize the overall performance of the tool (Duwe & Kim, 2016; Georgiou, 2019; Hamilton et al., 2015). Of course, when considering modifications to a risk instrument, it is imperative to evaluate if the revised version is capable of outperforming the original (Cunningham & Sorensen, 2006; Duwe, 2019). (Labrecque 2022, p. 49)*

For these reasons, it is important that correctional authorities engage in defensible factor selection for classification tools and assign inmates to appropriate custody levels based on empirical estimates of misconduct risk instead of relying on casual intuitions about how specific factors should be weighted or allowing the subjectivity of correctional officers' personal preferences to structure custody placement decisions.

When designing a classification tool intended to predict misconduct, it is also important to consider whether the goal of the tool is to predict general misconduct (GM) or serious violent misconduct (SVM) since they have different prevalence rates within the prison population and different predictor sets. It is also important to understand how much misclassification error a correctional authority is willing to accept and how to define appropriate benchmarks for tolerable classification error. For example, if a classification tool predicts GM accurately 80% of the time and SVM accurately 60% of the time, it may still be useful if the desired use of the tool for correctional authorities is to reduce overall misconduct and not just SVM.

Overview of Phase III Study Goals

ISR's methodology for Phase III is based on recommendations developed by the National Institute of Corrections (NIC). Since the completion of ISR's Phase I report, the NMCD has undergone important programmatic changes, including a shift towards special population management, policy adjustments regarding discipline, and conversion of private prisons to state-operated facilities. Furthermore, the COVID-19 pandemic that began in March 2020 impacted NMCD's operations.

In this section, we update our Phase I review of NMCD policies affecting classification, including discipline and special management policies, as these policies intersect with the classification process in ways which impact the scope of our validation results. We provide a list of review policies in Table 1.

Table 1.
Reviewed Policies

Policy Number	Policy Description
080100	Institutional Classification, Inmate Risk Assessment and Central Office Classification
081200	Institutional Classification and Risk Assessment
080200	Good Time Deduction (MGT)
080400	Earned Meritorious Deduction (EMD)
080600	Meritorious Deductions
082700	Restoration into Population Program (RPP)
082800	Lump Sum Awards (LSA)
083000	Reentry Planning & Transition Process
083100	Unit Management

Classification Components

A properly implemented objective prison classification system should identify the level of risk presented by newly admitted inmates based on valid and reliable information (Austin & Hardyman, 2004). To ensure that the NMCD is correctly assessing and housing inmates, it was necessary to review both the classification system essential components as well as the classification major components that were established in Phase I. Table 2 highlights these components and has been updated from Phase I to reflect current policy and practice. Additionally, Table 3 updates the essential classification components reviewed in Phase I.

Table 2.
Classification System Essential Components

Component	Covered in Policy	Practice
Mission Statement	Yes	Yes
Classification goals and objectives	Yes	Yes
Dedicated classification unit and staff	Yes	Yes
Centralized control over prisoner transfers and housing decisions	Yes	Yes
Accurate prisoner data	Yes	Yes
Automated data systems	Yes	Yes
Continuous monitoring and process evaluation	Yes	Yes
Impact evaluation	N/A	No

Table 3.
Classification Process Major Components

Component	Covered in Policy	Practice
Inmate identification	X	Recording basic personal data and information is gathered and entered into Criminal Management Information System (CMIS)
Medical health screen	X	New inmates coming through RDC go through a medical health screening upon admission to NMCD
Physical examination	X	New inmates coming through RDC go through a physical health screen upon admission to NMCD
DNA testing	X	Completion of DNA sampling occurs on applicable inmates
Social history	X	The social history of an inmate is compiled upon admission to NMCD and put into the Admission summary report
Custody level	X	Objective assessment of each inmate's risk of violence of escape is Administered
Internal classification	X	Each inmate is assigned to a Classification Officer and internal classification is completed within four weeks upon arrival at RDC
Prisoner separation	X	Based on inmates request and/or an in-depth review of the facts and documentation indicating that, if placed in GP, the inmate would be in jeopardy of serious bodily harm
Gang membership	X	Inmates admitted at RDC go through Security Threat Intelligence Unit (STIU) screening to identify gang affiliation + risk
Victim notification	X	The AODA Victim Notification Coordinator is responsible for reporting offender status to the statewide DA's offices who in turn notify registered crime victims
Academic achievement	X	A representative of the Education Bureau shall interview all inmates at RDC for a summary educational assessment
IQ tests		
Vocational aptitude	X	A representative of the Education Bureau shall interview all inmates at RDC for a summary vocational assessment
Substance abuse testing	X	Simple question on initial custody scoring form is asked about drug abuse in the last 10 years
Psychological testing	X	New inmates coming through RDC go through a behavioral health screening upon admission

Classification System Update

We observed the classification process to verify policy adherence. In Phase I, policy compliance was not assessed by the evaluation team. However, by adding observations in Phase III, we gained better insight into how policy is implemented across facilities and components of policy adherence. In Table 4 and Table 5, we present differences in the initial classification processes at CNMCF and WNMCF and reclassification processes at various facilities. These tables enhance our understanding of the NMCD classification system. Table 4 includes data from two observers on initial classification components and staff completion. Table 5 includes staff tasks for external classification and reclassification based on observations.

Table 4.
Internal Initial Classification Components Observed at RDCs by ISR Staff

Classification Components	Covered In Policy	Observed by ISR Staff
Review of Casework Materials	X	X
Pre-Institutional Assessment Information	X	X
Admission Summary	X	X
Initial Custody Scoring Form	X	X
Recommendation to Supervisor	X	X
Interview with Inmate	X	X
Prison Rape Elimination Act (PREA) Screening	X	X

Table 5.
External Classification/Reclassification Components Observed at Facilities by ISR Staff

External Classification/Reclassification Components	Covered in Policy	Observed
Mandatory	X	X
Review of all relevant paperwork	X	X
Updates in CMIS	X	X
Rescore	X	X
Recommendation to Manager	X	X
Review with Inmate (optional)	X	X
Discretionary	X	X
Review of Behavioral Infractions	X	X
Work/Program Assignments	X	X
Medical/Mental Health Needs	X	X
Rescore	X	X
Recommendation to Manager	X	X

NMCD Classification Focus Groups

Focus Group Themes Comparisons

We conducted a series of focus groups with classification and supervisory staff in Phase I. The focus groups resulted in 10 major themes from four focus groups. In Phase III, we conducted five focus groups and identified nine major themes. Table 6 presents the overlap and differences in themes between Phase I and Phase III's focus groups. We describe important variations in discussions later in the report.

Table 6.

A Comparison of Focus Group Themes in Phase I and III

Theme	Phase I	Percent	Phase III	Percent
Policy	X	18.3%	X	9.2%
Process	X	16.0%	X	14.8%
Mission Driven vs Levels	X	3.0%	--	0.0%
Scoring and Scoring Forms	X	26.1%	X	20.0%
Overrides (Mandatory/Discretionary)	X	8.7%	X	9.5%
Issues/Challenges	X	8.1%	X	16.1%
Training	X	4.2%	--	0.0%
PREA	X	1.2%	--	0.0%
COMPAS	X	9.9%	--	0.0%
Recommendations	X	4.5%	--	0.0%
Sex Specific Scoring Forms	--	0.0%	X	4.0%
Purpose	--	0.0%	X	10.1%
Programming and Services	--	0.0%	X	6.0%
Miscellaneous	--	0.0%	X	10.3%
Total		100.0%		100.0%

Between November 2021 and April 2022, ISR conducted five focus groups with 46 NMCD classification and supervisory staff to assess their perspective on NMCD's classification scoring forms, process, and procedures. The focus groups included staff from CNMCF and PNM, staff from SNMCF, staff from LCCF, supervisors at NMCD's Central Office, and staff from WNMCF and SCC and lasted approximately 90 minutes each. We used focus groups to add depth to our review of NMCD policy and classification data (Gibbs 1997; Stewart & Shamdasani, 2014), gathering information regarding participants' thoughts about the classification scoring form, how policy interacts with the classification process, improvements that can be made to the scoring form, and how the form fits into the broader classification process. The focus groups assessed staff beliefs about whether classification policy and practice align, how staff felt about the use of the different forms and the inclusion of specific factors on the forms, the use of mandatory and discretionary overrides, and to solicit recommendations for process improvements. Table 7 describes the focus group participants.

Table 7.

Characteristics of Focus Group Participants

Demographics	Count	Percentage
Classification Officers	35	76.0%
Unit Managers and Administration	10	22.0%
Completed College or Master's Degree	22	48.0%
Average Age	41	--
Sex: Female	35	76.0%
Average Years in Corrections	13	--
Average Years in NMCD	8	--
Average Years in Classification	6	--

N = 46

Focus Group Themes from Phase III

ISR's five focus groups in Phase III yielded general findings. Thematic analysis extracted nine focus group themes, which Table 8 reports on. These themes focus on the policy, process, training, custody levels, programming and services, and challenges as they relate to the scoring forms and factors.

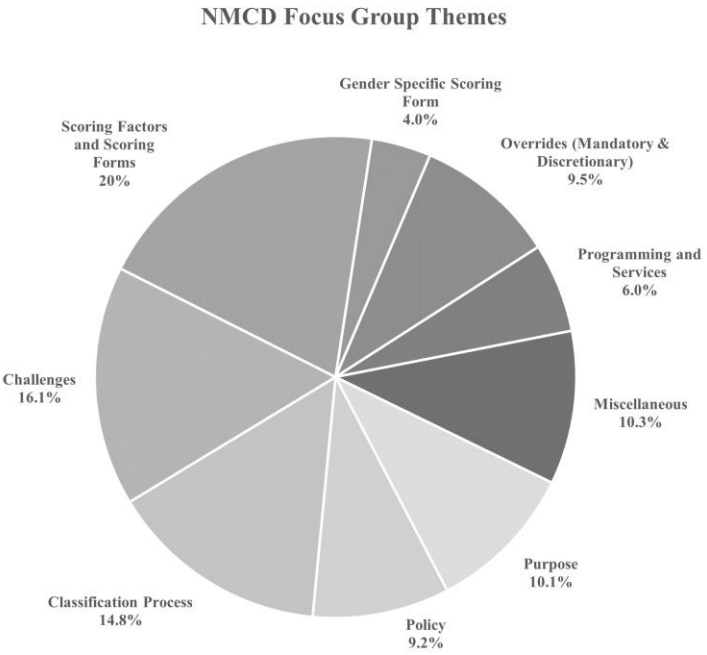
Table 8.
Focus Group Themes and Definitions

Theme	Definition
Scoring Forms and Scoring Factors	Includes the classification and reclassification scoring forms, as well as the scoring factors on both forms
Classification Process	From the point of initial custody by NMCD through external classification at the RDC, as well as the process of internally classifying and reclassifying inmates at assigned facilities
Challenges	Common difficulties expressed among participants
Overrides (Mandatory and Discretionary)	Factors that are listed on the scoring forms
Policy	NMCD policy 080100 and other policies it interacts with
Programming and Services	The options for inmates available at facilities that include medical and mental health needs, as well as different skill-based classes and resources to which inmates have access
Sex Specific Scoring Form	Scoring forms that are specific to male or female populations
Purpose	The purpose of classification in institutions
Miscellaneous	Includes various diverse topics of conversation that do not fall under any one particular theme

Staff in the focus groups centered their policy discussions on occasional conflicts between current classification policy and disciplinary policy. Staff discussed the classification process at each facility, from initial classification at RDC to external classification and reclassification at other facilities. Participants provided feedback on whether and how scoring tool factors should be adjusted, whether medical and mental health restrictions should be factors or overrides, and how these changes could be beneficial. Discussions on overrides focused on the appropriate use of mandatory versus discretionary overrides and the overrides that were most used. Challenges to classification such as training, directives conflicting with policy, and communication barriers were also frequent themes. When discussing programming and services, participants often mentioned the shortage of resources across all facilities and how this ultimately impacted inmate placement.

Figure 1 visually presents the major themes as a percentage of the total time spent discussing them in all five focus groups. Scoring and the scoring form was the most frequently discussed topic, accounting for 20% of the overall time. Challenges (16.1%), Classification Process (14.8%), Purpose of Classification (10.1%), Overrides (9.5%), and Policy (9.2%) were also significant themes. Programming and Services (6.0%), Sex Specific Scoring Forms (4.0%), and Miscellaneous (10.3%) contributed to the remaining time.

Figure 1.
Themes from Phase III of NMCD Focus Groups



In what follows, we provide additional context for some of our focus group discussions and provide a sample of representative quotes from correctional officer staff and supervisors.

Purpose

Participants in the five focus groups discussed the purpose of classification in corrections. Although all participants agreed that classification was crucial for maintaining the safety of inmates, staff, and the public, they had different opinions on the purposes of the tool, which demonstrated that the tool serves multiple purposes. For example, the following quotes identify three different perceptions of the tools' purposes, including institutional safety, post-release recidivism reduction, and access to programming and appropriate services.

> *I'll say it's for overall the safety of the inmates, the staff, the community to make sure everyone's placed where they should be in case you have somebody with escape history. You don't want to put them* (those with an escape history*) at a Level 1 institute for work reasons.*

> *. . . I feel like it's to help them to keep from not coming back. There's a lot of things that we don't know about their life. We're . . . just trying to help them to know that there is help out there, that [we] can help them to move forward instead of doing the same thing over and over.*

> *. . . To make sure that they are at the right facility where they can get the appropriate care. . .*

The last quote about the tool serving the purpose of assisting inmates secure access to appropriate services reveals a potential confusion between the goals of internal and external classification systems and a

potential tension between the primary purpose of the tool (e.g., reducing institutional violence) and a secondary consideration related to service availability.

Policy

Participants generally agreed that the current classification policy serves its purpose and is helpful as a general guide, but some expressed dissatisfaction with the fact that the impact of policy depends on the population being housed. Although participants acknowledged the effectiveness of the most current version of the policy, they highlighted the need for improvement, particularly in how the policy addresses different scoring factors and conflicts with the disciplinary policy. Several groups discussed the shortcomings of the new policy and forms in identifying repeat disciplinary offenders which they considered the most challenging aspect of the current policy. This issue is further explored in the *Overrides* subsection.

> *. . . They get parts of it and use it the way it needs to be used for each classification or facility. . . They try to keep within the confines of the classification policy. . . They do change it up for the different areas of the department. It's not broken up into special management programs. It's not broken up into general population. It's not broken up in areas like that. It's one universal policy that they get and use it for all of them and then change it up how they need for each area that they need.*

> *I appreciate that the policy is reviewed annually. . . I don't have any complaints about the policy right now. I do believe that there is a need for that annual review and constant change. . . with people, there's lots of factors involved that we can't predetermine that are going to take place. . . I don't think the policy change kind of fixes some of the stuff we were talking about with the module CMIS and how it produces its information . . . having some of those changes in policy and actually being affected in our management system accurately would assist.*

> *A good policy's there to check to everybody around but . . . because of the dynamics within corrections, if they follow policies and procedures in the classification aspects, you'd have less movement.*

Classification Process

The focus group members demonstrated knowledge of the classification process and provided detailed descriptions of it. They reported being aware of many components relevant to the initial classification process, including criminal record, medical and mental health, PREA, and STIU. Despite being understaffed, participants expressed satisfaction with both the amount and quality of work they were able to produce. The classification process at facilities was the second most discussed topic (15%), following the scoring forms and factors (20%).

Broad Challenges

Classification officers expressed challenges, including the lack of formal training and ongoing training. They suggested having specific policy and classification training, annual refresher training, and creation of a formal training manual.

> *. . . There's so much information to be learned that . . . somebody coming off the street or even coming in needs to be taught. And constantly there always has to be, refreshers on it.*

> *. . . There's really no formal training . . . they have classification 101. Mainly what that one does is you go in and they go over policy, but it's not full case policy. They kind of go through it real fast and say this is what you're looking for. There's no real formal training of how to do classification. . .*

Participants also highlighted the need for better communication among staff across facilities, particularly regarding policy changes and directives.

> *We've got to figure out how to work together. Every facility has to work together.*

Some staff expressed concerns about the confusion between policy and directives and the need for clear policies and procedures.

> *Daily directives, policy, and we're like–it's like as a collective you never asked me. You just kind of went, 'All right, because I do it this way, everybody should do it this way.*

Participants also discussed the challenges of CMIS, the electronic system used to score and manage inmates, and suggested that the system automatically track mandatory overrides to streamline the scoring process and reduce discrepancies among classification staff.

> *And that's a problem with the system as a whole. It can be a training issue. . . mandatory overrides should never be manually entered. It's something the system should identify, which hopefully with OMNI, some of these mandatory overrides will be picked up and they'll go into effect because when you're. . . referring inmates to lower custody and you have to look at their medical, mental health, and all that, those should be automatic triggers in there [so] you can't classify an inmate lower than what he is cleared for medically, mentally, or even based on the mandatories.*

Scoring Factors and Scoring Forms

Most participants agreed that the scoring form requires minor updates. They suggested separating mental health and medical restrictions on the form and adding them as scoring factors versus mandatory overrides, which could be helpful.

> *. . . On the scoring form, on the second page, it'll have level 1 restrictions listed as medical or mental health holds for whatever reason, and the unit manager [said]. . . 'I think those need to be separated,' so that we can determine which one it is versus trying to guess.*

> *. . . I think I already expressed that I do feel that it's wrong. . . The guy couldn't go to that level because he was on insulin, right? And you're holding that against someone–to hold him at a higher custody level when he's lower but he just has a medical [restriction].*

Institutional adjustment for priors and gang affiliations were also discussed during the focus group sessions. Participants spoke about the impact of the scoring form on the classification system, particularly in terms of scoring related to disciplinary infractions.

> *. . . Sometimes they go through the RPP program. And those gang affiliations still stay at the bottom. . . if they're not affiliated, then why are they staying at the bottom? So they're still getting points for it. . . because it still has a gang affiliation. So that's how I feel about the gang affiliation. Anytime the inmates get new tattoos, if they have something that's gang-affiliated, then they should be reporting it so that way [it] can be put on their score sheet. . .*

Most classification officers mentioned that the scoring seems to be more relaxed than the previous scoring tool. Some participants explained what happens when an inmate picks up a disciplinary infraction, indicating that in the current policy, an infraction can only count against an inmate once, and the behavior can be repeated without an inmate picking up more points.

Focus group participants also discussed the fact that incidents stay on an inmate's record for 10 years. They discussed how incidents impact the classification score and that old incidents could potentially keep an inmate at a higher security level when no other disciplinary reports had occurred in the recent past. Participants felt that the 10-year threshold could be lowered.

> *. . . In some cases, whether or not it can be perceived as violent. . . 10 years is a long time to score someone. I don't mind taking into consideration factors like that when reviewing someone for placement at lower custody level, but you do have guys that [have] pretty significant stretches of clear conduct together that are still being penalized or held in a 4 for reasons like that. In some cases though, it does fit the individual where. . . they're housed in a level 4 custody level and can remain report-free and they get to a level 3, and then just like that, they're back on that track of violence against other inmates, violence against staff, and they're back at the PBMP, and there goes their misconduct history all over again. I think it probably could be addressed another way. And I think that giving someone a light at the end of the tunnel is a good motivation for maintaining clear conduct. 'If I don't have to hit the 10-year benchmark, I can do the 5-year, and then maybe before I leave, I can hit a level 2.' Where now, 'It doesn't matter what I do. I'm stuck in a level 4. I might as well continue to act up and be unruly.'. . .*

Gender Specific Scoring Form

During Phase I focus groups, classification officers who worked at female facilities advocated for a separate tool for females. Although women have unique needs and participants in Phase III's focus group for female facilities agreed there are differences in behaviors and tendencies in the women inmates, most classification officers felt that a separate form for females was not necessary.

> *. . .I thought to a certain extent that it shouldn't matter whether you're a female or male, because as far as being scored. . .what scoring really pertains to you being female or male?. . . A female's obviously capable of what a male's capable of doing, so there's no override that's gender-specific or that can be changed based on gender.*

> *. . . I think it should just be straight across the board. . . a crime is a crime. And it should be no different if you're a man or [woman].*

Overrides (Mandatory versus Discretionary)

Participants shared diverse experiences, perspectives, and opinions regarding mandatory and discretionary overrides. They briefly discussed mandatory overrides and highlighted the need for CMIS to automatically score mandatory overrides to facilitate the classification process. Disciplinary infractions were the most common reason for discretionary overrides and were widely discussed in all focus group sessions. Validating the tool was considered useful to determine the frequency and timing of these overrides. Participants also discussed the differences between mandatory and discretionary overrides and suggested that some override factors could be redefined. While certain mandatory overrides, such as mental health or medical reasons and repeat disciplinary offenders, were considered necessary, the issue of repeat disciplinary offenders who engage in the same type of infractions was highlighted by the staff.

Existing NMCD disciplinary policy allows counting disciplinary convictions once per class within a 24-month period, which may result in inmates continuing to engage in the same behavior. Many participants expressed their desire to change this and suggested utilizing overrides, but since it rarely affects an inmates' points and puts them in a higher custody level, reaching out to the Central Bureau of Classification (CBC) is a more common approach to get repeat disciplinary offenders housed at a higher level. Participants discussed inconsistencies between the classification policy and the disciplinary policy, noting that this inconsistency caused confusion among inmates.

> *. . . We [have] interacted with these inmates so much that we know these inmates. We know that they're a nuisance. We know that they're problem children. We know that they need that management control, [yet] our scoring system is saying, "No, he meets level 3 criteria," but we don't feel comfortable sending him to a different level.*

> *I think that [on] the disciplinary side. . . they should be able to score them multiple times on the same report not just one time every year or every two years, either from that minor two years for a major. . . They should be able to be scored based on the amount of times that they pick up reports.*

> *If they're already maxed out on that point level where they only get 11 points, [and] they pick up another report it's not going to matter because those points are not going to change. If they get clear conduct for a year, then their points are going to drop even though they're still picking up those same reports. So it's still going to drop. . . it's not helping them.*

The topics of a less flexible disciplinary policy and the overlap between the classification policy were mentioned multiple times in all focus groups. Participants highlighted the difficulty in explaining this overlap to inmates and suggested the need for clear policies and procedures that align with the department's expectations.

> *. . . I know some changes were made last year, and some of the changes they made actually hurt the facility more than they hurt. . . the inmate, especially when it comes to misconduct reports.*

> *Some of the major changes they made were if you get the drug misconduct report[s] from scoring 4 or 5 points in the misconduct section to scoring 1 or nothing. . . so that's really affected an inmate, especially at a level 2 facility if he picks up a misconduct report because the next time he's reclassified, he can still end up ultimately going to a level 1, especially if they don't catch'the mandatory override of the 6 months for having received a drug-related misconduct report. . . it allows him to go a lower custody even though they had recent drug-related misconduct reports. . . Some of the other things that were a little lower-level severity, they bumped up to higher-level severities. . . some of the changes they made . . . weren't beneficial to the proper housing for inmates. . .*

Programming and Services

Participants in all focus groups expressed a need for more programming and services at all facilities. They noted that if more programming and services were available at all custody levels, the NMCD would not need to use mandatory overrides for mental health and medical restrictions which currently override inmates to higher custody levels because necessary services are not available in some lower level facilities. The lack of programming and services was discussed as a factor related to mandatory overrides.

> *. . . Services shouldn't be dependent on their security level.*

> *. . . The main issue [is] statewide, we have an issue with behavioral health providers, especially there at Central. And so the level 2 is now a 1, 2, geriatrics. . . They're not getting the programming they need to address their drug issues. . . It makes it worse at the level 1 facility because we don't have anybody there now to address those needs.*

> *More options for the inmates would help . . . in the scoring aspect because it's like I tell my guys. . . you need to start planning for your future. Learn those tools that the majority of us. . . learn how to do a checkbook and all that stuff. . . The majority of the inmates are people that [have] dynamics of single-family homes, poor education, poor, don't have that mindset of how to think outside the box. . .*

Furthermore, all facilities reported a shortage in the programs offered, with some having waitlists, and others being impacted by Covid-19 and not starting up again. In the female focus group, it was evident that there was a greater need for programming specific to females as they have limited access to facilities with a diverse training portfolio. Classification staff generally agreed that inmates should have access to tools that can help them gain life skills and programming can deter inmates from negative behavior and disciplinary infractions.

> *. . . 'They don't have anything. They have very few programs that females can do, very few places they can [transfer] to.*

Focus Group Conclusions

Classification staff provided insights on a range of topics during the focus groups, including scoring and scoring forms, the challenges they face during the classification process, the purpose of classification, overrides, policy, programming and services, gender-specific scoring forms, and other miscellaneous issues. These discussions were crucial in helping us understand the impact of policy on classification staff

and identify areas that could be modified to facilitate classification processes. Although the primary aim of the current report is to validate the existing NMCD classification tools, the focus groups allowed us to engage with staff and gather their feedback on their experiences with the scoring forms and proposed changes actively. The conversations provided contextual insights on the scoring tools that a quantitative analysis of the tools alone could not provide. The classification staff's real-life use of the tools helped us gain a better understanding of their perception of the scoring tools' strengths and weaknesses, which – in tandem with the quantitative validation study results – helps inform our recommendations for tool revisions going forward.

Observations

We conducted observations with NMCD classification staff at different facilities between November 2021 and February 2022 to collect detail on the use of the current NMCD classification scoring forms and the classification process's relation to policy. We observed the process of scoring the initial classification and reclassification tools, aiming to provide additional context and depth to our review of NMCD policy and classification data. Conducting observations enables us to view and explore intimate interactions among naturalistic environments (Fox 1998). The nature of the observations provided rich data that supplemented the analysis of policy and classification forms, providing us with real-life experiences that could not be obtained from document review or surveys.

Informed Conversations

During observations, we held informal structured conversations with staff members. Staff members discussed various topics, including how they update initial scoring form factors, their perceptions of custody level scales, the relationship between special management status and custody level, who updates CMIS, custody override factors, reclassification scoring form factors, and their general philosophy regarding inmates.

Observational Analysis

To confirm adherence to policy, we developed a checklist that aligned with the tasks listed in policy. Additionally, we engaged in informal conversations with classification staff regarding the process, providing further insights into their approach to classification. By using these different methods, we gathered detailed data on the initial classification at both RDCs, as well as internal classifications and reclassifications. This allowed us to organize and compare data to ensure that the process and policy align. Table 9 displays the major components observed during the external classification process at the RDCs. This approach provided valuable information on how staff members approach their work and allowed for a more holistic assessment of the classification process.

Table 9.
RDC External Classification Major Components

External Classification Components in NMCD Policy 080100	Observed at CNMCF	Observed at WNMCF
Review of Casework Material	X	X
Gather Pre-institutional Assessment Information	X	X
Interview Inmate		X
Develop Admission Summary	X	X
Complete Initial Scoring Form	X	X
PREA Screening		X
Recommendation to RDC Manager	X	X

During conversations with CNMCF staff members, we discovered that due to Covid-19, the facility was not conducting inmate interviews. Instead, inmates were given either a 'Long Form' or 'Short Form' to complete to provide information about their needs and institutional history. Classification staff used this information to complete the initial scoring form. Observers were only able to witness classification staff inputting information into the system based on the completed forms, and no inmate interviews were conducted during the observation period. Additionally, the PREA screening process was not observed at CNMCF. The process for internal classifications and reclassifications differed from the initial classification process. Table 10 provides a description of the major components observed at CNMCF, SNMCF, LCCF, and WNMCF. Although the lack of inmate interviews limited the scope of observations, the use of Long and Short Forms offers insight into how classification staff adapt to challenging situations, which may be beneficial in developing new policies and practices in the future.

Table 10.
Internal Classification and Reclassification Major Components

Internal Classification and Reclassification Components in NMCD Policy 080100	Observed
Review Relevant Paperwork	X
Disciplinary Reports (Discretionary Only)	X
Work & Program Assignments	X
Medical & Mental Health Needs	X
Updates in CMIS	X
Rescore	X
Recommendation to Unit Manager	X

Data Analysis – Validation of the Initial and Reclassification Tools

In this section, we evaluate the predictive validity of the NMCD's initial and reclassification tools. We assess the tools' predictive performance by sex and misconduct type [i.e., general misconduct (GM) versus serious violent misconduct (SVM)] and determine whether there were differences in misconduct by custody level. Additionally, we explore variation in misconduct within and across custody levels based on whether an inmate was overridden.

Inclusion and Exclusion Criteria for the Validation Analysis

We collected data on classifications and disciplinary events for all inmates admitted to NMCD between January 1, 2015, and December 31, 2021. We obtained records for 12,455 unique inmates with initial intake classification and 13,268 unique inmates with reclassification. To create the validation sample, we excluded records for inmates who were released within the six months following their classification, duplicated records, and records for inmates overridden into custody levels that did not match the custody level generated by the classification tool. These exclusions reduced the validation sample to 4,915 unique inmates with initial classification scores and an equivalent number of total classification records, as well as 7,904 unique inmates with reclassification scores and 29,331 reclassification records. Consequently, we were able to use 66% fewer initial classification records and 57% fewer reclassification records for validation[2].

Measuring Misconduct

We used two different measures to assess institutional misconduct. The first measure includes any disciplinary violations that an inmate was found guilty of within six months after being classified. The second measure includes only serious violent disciplinary violations that an inmate was found guilty of within six months after being classified. We chose a six-month follow-up period since the NMCD reclassifies inmates every six months, except for Level I inmates and those with a Special Management designation, who are reclassified annually. We analyzed the data using logistic regression to predict whether an inmate would engage in any misconduct within six months and negative binomial models to predict the total number of misconduct incidents. Within six months of initial classification, 46% of inmates engaged in GM at least once, while only 6% engaged in SVM. Within six months of reclassification, 53% of inmates engaged in GM at least once, whereas only 7% engaged in SVM. Table 11 provides a demographic breakdown of the validation sample.

Table 11.

Descriptive Characteristics of Unique Inmates Within the Validation Sample

	Initial Classification	**Reclassification**
Sex: Male	86% (4,241)	87% (6,900)
Sex: Female	14% (674)	13% (1,004)
Age: 18-21	6% (279)	4% (315)
Age: 22-25	16% (760)	16% (1,281)
Age: 26-34	38% (1,852)	41% (3,274)
Age: 35-44	28% (1,363)	26% (2,027)
Age: 45+	13% (661)	13% (1,007)
Race-Ethnicity: Hispanic	62% (3,025)	62% (4,922)
Race-Ethnicity: White	21% (1,052)	22% (1,703)
Race-Ethnicity: Native American	8% (377)	7% (530)
Race-Ethnicity: Black	6% (296)	6% (490)

[2] Approximately 52% of initial classification records and 42% of reclassification records were ineligible for validation due to overrides. Of all overrides, 93% were mandatory and 7% were discretionary.

Race-Ethnicity: Other	3% (165)	3% (259)
Total (*n*)	4,915	7,904

Within the validation sample for the initial classification tool (*n* = 4,915), the average age at the time of assessment was 34 (SD = 9.5, range = 18-79), and 86% were men. 62% of the validation sample were Hispanic, 21% were White, 8% Native American, 6% Black, and the remaining 3% were a different race or ethnicity classification. Within the validation sample for the reclassification tool (*n* = 7,904), the average age at the time of assessment was 34 (SD = 9.3, range = 18-81), and 87% were men. 62% were White (Hispanic), 22% White (Non-Hispanic), 7% Native American, 6% Black, and the remaining 3% were a different race or ethnicity classification. On demographic characteristics, the validation samples for the initial and reclassification tool did not vary in statistically meaningful ways[3].

44% of inmates in the validation sample participated in prison programming within six months of their initial classification (*n* = 2,155), and 30% participated within six months of their reclassification (*n* = 2,378). Tables 12 and 14 outline the programming type inmates most participated in by classification level, and Tables 13 and 15 show program participation by custody level with educational programming being the most common (e.g., 24% and 17% of inmates participated in Adult Basic Education programming for initial and reclassification samples, respectively). A chi-square test of independence indicated that there was a significant negative relationship between custody level and participation in prison programming within both the initial and reclassification samples, [Initial Classification: X^2 (3, N = 4,915) = 228.1, $p < 0.01$; Reclassification: X^2 (3, N =7,904) = 120.13, $p < 0.01$] suggesting that, in general, as custody levels increased, programming entry decreased. It is possible that program participation is higher when considered across a longer time frame than the six months before or after classification and that the present measurement of program participation is overly blunt as it only captures program entry and does not reflect either (1) the duration of time an inmate spent within a program or (2) whether an inmate completed the program, the latter of which may exert more of a meaningful effect on institutional misconduct than program-entry alone (for a more detailed discussion on this point, see Pompoco et al., 2017).

Table 12.

The Most Common Prison Programming of Inmates in Six Months Following Initial Classification

Program Type	**Frequency**	**Count**
Educational Programming (e.g., Adult Basic Education)	24%	1,196
Addiction Programming (e.g., ASB OP Treatment)	12%	593
Special Management Programming (e.g., Reentry PPD Men's/Women's Recovery)	5%	236
Cognitive and Life Skills Programming (e.g., Moral Reconation Therapy; Charting a New Course)	2%	100

[3] We assessed this statistically by predicting classification type (0 = Initial Classification; 1 = Reclassification) from a feature set inclusive of all demographic variables using a logistic regression. Results of the logistic regression indicate there were not statistically significant differences in the demographic profiles of inmates based on whether they were part of the initial or reclassification validation sample.

Table 13.

Program Participation by Level in Six Months Following Initial Classification (n = 4,915)

Score-Based Level	Frequency	Count
Level I	49%	395
Level II	49%	1,571
Level III	23%	185
Level IV	5%	4

Table 14.

The Most Common Prison Programming of Inmates in Six Months Preceding and Following Reclassification

Program Type	Frequency	Count
Educational Programming (e.g., Adult Basic Education)	17%	1,317
Addiction Programming (e.g., ASB OP Treatment)	8%	631
Special Management Programming (e.g., Reentry PPD Men's/Women's Recovery)	3%	264
Cognitive and Life Skills Programming (e.g., Moral Reconation Therapy; Charting a New Course)	2%	132

Table 15.

Program Participation Custody Level in Six Months Preceding and Following Reclassification

Score-Based Level	Frequency	Count
Level I	35%	582
Level II	33%	1,140
Level III	25%	582
Level IV	15%	74

We present additional descriptive characteristics of the joint validation sample – inclusive of the initial and reclassification validation samples – in Tables 16-18.

Table 16.

The Five Most Common Charges for Inmates Based on Classification Status

	Initial Classification	Reclassification
Possession of a Controlled Substance – Felony	33% (*n* = 1,600)	28% (*n* = 2,246)
Trafficking Controlled Substance	16% (*n* = 805)	14% (*n* = 1,143)
Burglary of Vehicle	10% (*n* = 507)	12% (*n* = 951)
Receipt/Transport/Possession of Firearms as a Felon	10% (*n* = 476)	8% (*n* = 636)
Resisting Arrest/Evading an Officer	9% (*n* = 459)	--
Residential Burglary	--	10% (*n* = 826)

Table 17.
Historical Gang Affiliation by Known Gang Type

	Initial Classification	**Reclassification**
No Prior Gang Affiliation	61% (*n* = 3,238)	58% (*n* = 5,022)
Street Gang	39% (*n* = 2,046)	41% (*n* = 3,557)
Prison Gang	0% (*n* = 18)	1% (*n* = 106)

Table 18.
Historically Most Common Gang Affiliations by Tool

	Initial Classification	**Reclassification**
Burquenos	5% (*n* = 274)	7% (*n* = 526)
Cruces Boys	3% (*n* = 150)	3% (*n* = 252)
Westside Locos	2% (*n* = 89)	2% (*n* = 170)
Southside Locos	2% (*n* = 87)	2% (*n* = 137)
Crazy Town	1% (*n* = 73)	1% (*n* = 109)

From Table 16, the most common types of charges for which inmates were incarcerated included possession of a controlled substance – felony (30%), trafficking of a controlled substance (15%), or burglary of a vehicle (11%). 70% of inmates received a sentence of less than two years. 22% were incarcerated for at least one violent offense. 30% of inmates were classified as having a suspected or validated gang affiliation in their history, with suspected street gangs accounting for 95% of overall gang affiliation. Additionally, 40% of inmates had an associated mental health flag during their incarceration. 3% of inmates were classified as having a severe or extreme mental illness episode within a six-month window of their initial classification or reclassification. Together, these descriptive findings highlight the prevalence of short-term sentences, non-violent offenses, gang affiliation, and, conditionally, mental illness, among inmates in the sample.

Data Analysis – Different Ways of Measuring Predictive Validity

To assess predictive validity, researchers employ various statistical methods. One commonly used approach is logistic regression. In this study, we used logistic regression to examine the impact of the tool-generated custody level and other relevant predictors on misconduct and predicted the probability of an inmate's misconduct within the six months following classification. We then evaluated prediction quality by comparing the predicted outcome with the observed outcome using area under the curve (AUC) statistics across and within levels.

A higher AUC value indicates better performance of the classification tool in distinguishing between inmates who engaged in misconduct and those who did not. The AUC is calculated from receiver operating characteristic (ROC) analyses and ranges from 0 to 1. An AUC of 0.50 means that the classification tool performs no better than a random guess and cannot differentiate between inmates based on misconduct risk. However, different fields of study have different standards for AUC quality thresholds. In our study, we adopted the benchmarks proposed by Desmarais and Singh (2013) to evaluate the quality of the predictive performance of the two tools (see Table 19).

Table 19.
Desmarais and Singh (2013)'s AUC Criteria

Quality	**AUC**
Excellent	0.71 – 1.0
Good	0.64 – 0.71
Fair	0.55 – 0.63
Poor	< 0.55

AUC statistics are generally preferred over accuracy (ACC) statistics because they are not base-rate dependent and can be compared across studies. We focus on AUC as the primary metric for evaluating the predictive validity of the tools given their frequent use in the criminology literature for evaluating the predictive validity of risk assessments (Lowenkamp and Betchel 2007; Fazel et al., 2022; Desmarais et al., 2021). However, we do want to highlight that recent research by Kwegyir-Aggrey et al., (2023) raises concerns about the use of the AUC as a primary metric for evaluating predictive validity specifically in contexts where multi-level (versus binary) classifiers are used. Because of this, we caution against making definitive claims about the tools' predictive validity by fixating on AUC results and note that the provided AUC estimates should be viewed tentatively for reasons more fully articulated by Kwegyir-Aggrey et al., (2023).

A Descriptive Overview of the NMCD Validation Sample

Table 20 provides a descriptive overview of the custody levels by classification tool and sex within the validation sample. Most inmates in the initial tool sample scored at Level II (66%), while in the reclassification validation sample, 45% of male classification records and 39% of female classification records were scored at Level III. We present visualizations of the distribution of tool scores by sex in Figure 2.

Table 20.
Descriptive Overview of NMCD Custody Levels within Validation Samples

	Initial (Male)	**Initial (Female)**	**Reclassification (Male)**	**Reclassification (Female)**
Level I	8% (*n* = 360)	66% (*n* = 475)	8% (*n* = 2,213)	24% (*n* = 565)
Level II	70% (*n* = 3,247)	32% (*n* = 228)	33% (*n* = 8,847)	39% (*n* = 933)
Level III	20% (*n* = 917)	2% (*n* = 17)	45% (*n* = 12,215)	31% (*n* = 733)
Level IV	2% (*n* = 90)	--	14% (*n* = 3,682)	6% (*n* = 143)
Total	4,614	720	26,957	2,374

Figure 2 reveals a limitation that only 17 female inmates were initially classified at Level III, and none at Level IV within the validation sample, resulting in significant noise being introduced into the predictive validity estimates. The low count of female inmates within Level III and IV makes the sample statistically underpowered at conventional thresholds, making it difficult to detect meaningful differences. To assess the scope of this problem, we conducted an *a priori* power analysis to determine the minimum sample size required to test the hypothesis of a positive relationship between custody levels and serious violent misconduct among female inmates in the initial validation sample. The results indicated that the cell size

of n = 16 for Level III and IV female inmates is too small to test hypotheses relating to the initial classification tool's generated custody levels and misconduct. To achieve 80% power for detecting a medium effect at a significance criterion of $\alpha = .05$, the required sample size for a multivariate general linear model with a balanced sample distribution across groups is n = 312. Because of this, we cannot confidently evaluate the predictive performance of the initial classification tool among female inmates for Level III and IV inmates.

Figure 2.

Scores on Initial and Reclassification Tools with Custody Level Scores by Sex and Tool

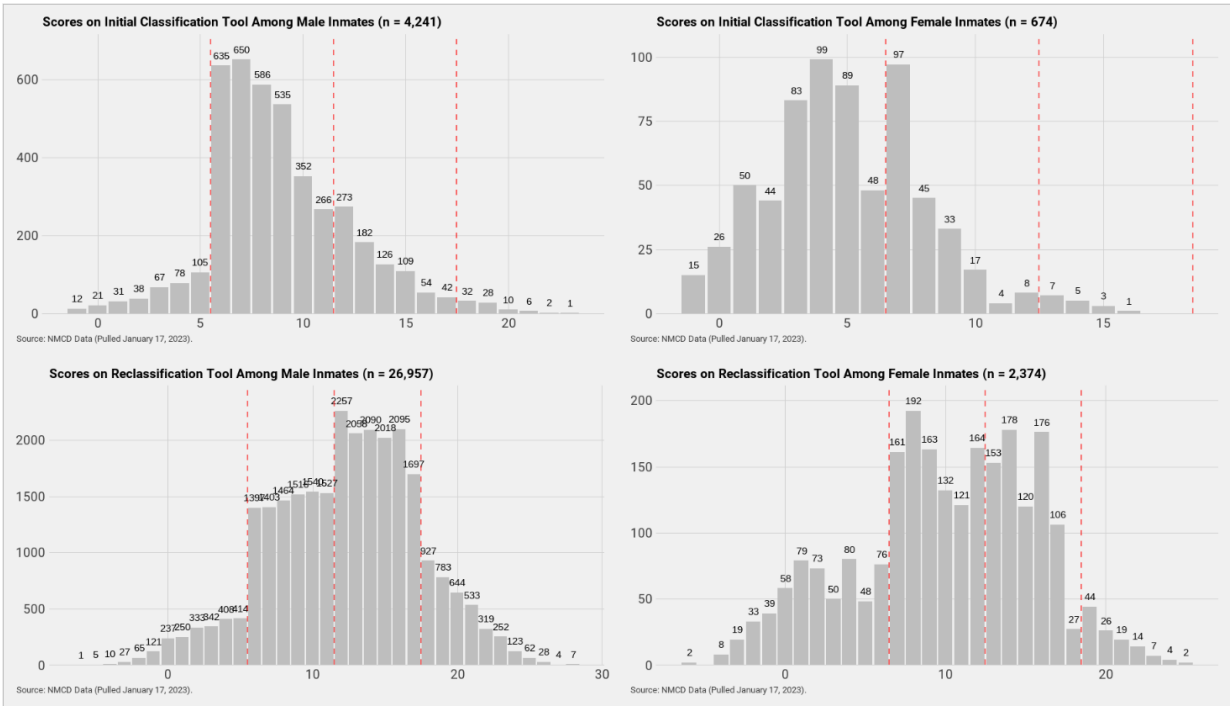


Table 21 displays the average number of infractions across the different custody levels by tool misconduct type, and sex.

Table 21.

Average Infraction Count by Level and Tool

	Initial (Male)	**Initial (Female)**	**Reclass (Male)**	**Reclass (Female)**
Level I – GM	0.63	0.87	0.51	0.43
Level II – GM	1.28	1.59	1.24	1.35
Level III – GM	1.61	2.59	1.44	2.28
Level IV – GM	1.48	--	1.07	3.07
Level I – SVM	0.02	0.02	0.02	0.02
Level II – SVM	0.06	0.07	0.05	0.03
Level III – SVM	0.07	0.12	0.07	0.08
Level IV – SVM	0.28	--	0.12	0.17

Table 21 shows that misconduct increases as custody level increases. This is true for male and female inmates, for the initial and reclassification tool, and for GM and SVM.

Table 22 displays the proportion of inmates in each custody level who engaged in misconduct in the six months following initial classification and reclassification.

Table 22.
Percent of Unique Inmates Engaging in Misconduct by Custody Level by Gender and Tool[4]

	Initial (Male)	**Initial (Female)**	**Reclass (Male)**	**Reclass (Female)**
Level I – GM	27% (*n* = 94)	34% (*n* = 154)	34% (*n* = 564)	30% (*n* = 115)
Level II – GM	45% (*n* = 1,477)	45% (*n* = 92)	62% (*n* = 2,123)	58% (*n* = 249)
Level III – GM	55% (*n* = 434)	50% (*n* = 8)	79% (*n* = 1,844)	71% (*n* = 117)
Level IV – GM	52% (*n* = 41)	N/A	80% (*n* = 398)	97% (*n* = 31)
Level I – SVM	2% (*n* = 7)	2% (*n* = 10)	4% (*n* = 70)	3% (*n* = 10)
Level II – SVM	6% (*n* = 170)	7% (*n* = 14)	12% (*n* = 409)	8% (*n* = 35)
Level III – SVM	7% (*n* = 55)	6% (*n* = 1)	20% (*n* = 467)	15% (*n* = 25)
Level IV – SVM	21% (*n* = 17)	N/A	31% (*n* = 155)	28% (*n* = 9)

GM was more prevalent than SVM in the six months following classification with an overall prevalence of 43% following initial classification (*n* = 2,302) and 44% following reclassification (*n* = 13,020), while there was only a 5% SVM prevalence following initial classification (*n* = 276) and a 6% SVM prevalence following reclassification (*n* = 1,627).

We conducted two-sample proportion t-tests to assess whether there were statistically significant differences in the proportion of inmates who engaged in GM and SVM between male and female inmates across different levels following classification. For the initial tool, results only revealed statistically significant differences in the proportion of male and female inmates engaging in GM with a higher proportion of females in Level I who engaged in GM than males. For the reclassification tool, results revealed statistically significant differences in the proportion of male and female inmates engaging in GM with higher proportions of male inmates engaging in GM than female inmates within Level I yet higher proportions of female inmates engaging in GM than male inmates in Level IV.

Figures 3-6 show the total number of inmates who engaged in institutional misconduct within six months of their classification categorized by the number of misconduct events.

[4] This table considers an inmate's entire incarcerated period. That is, if an inmate ever engaged in a misconduct event within six months following a classification event, they would be included as part of this percentage.

Figure 3.

Count of Inmates by Frequency of GM Events in Six Months Following Initial Classification (n = 5,334)

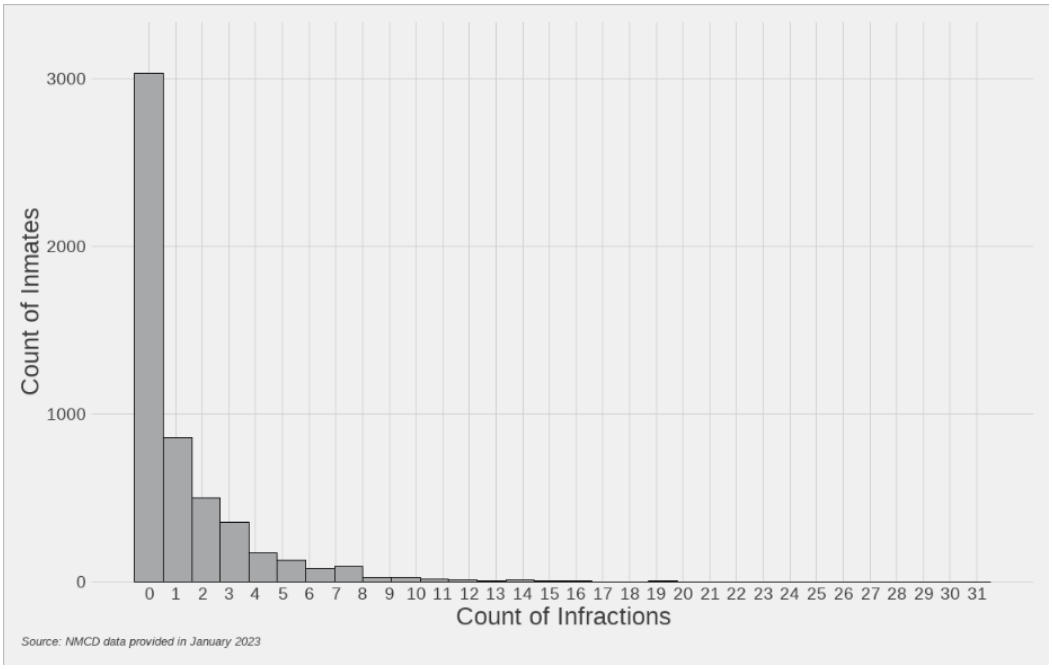


Figure 4.

Count of Inmates by Frequency of SVM Events in Six Months Following Initial Classification (n = 5,334)

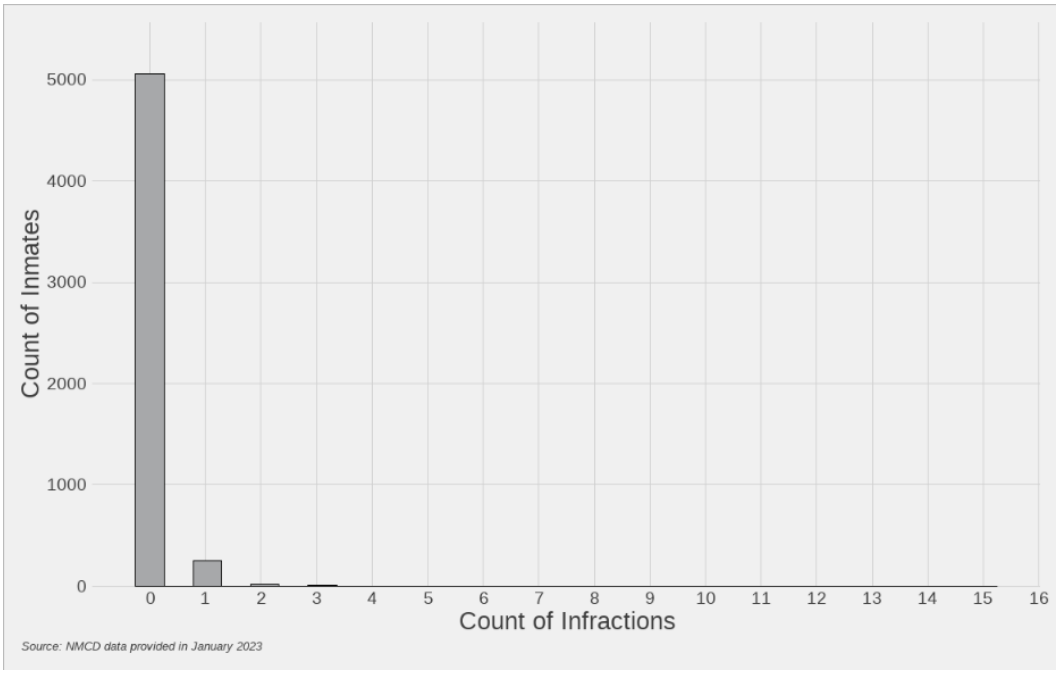


Figure 5.

Count of Inmates by Frequency of GM Events in Six Months Following Reclassification (n = 29,331)

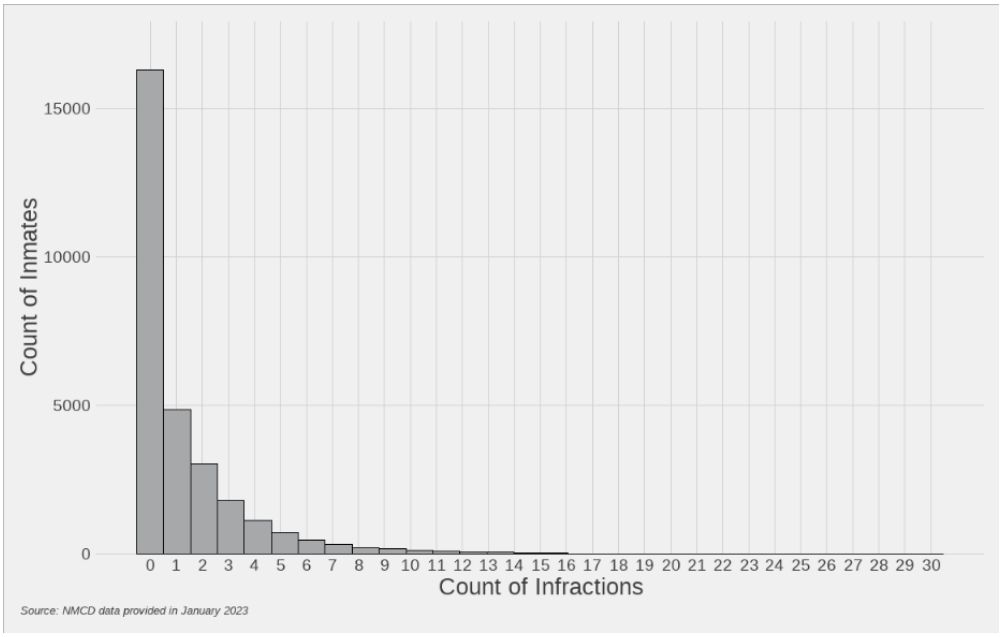
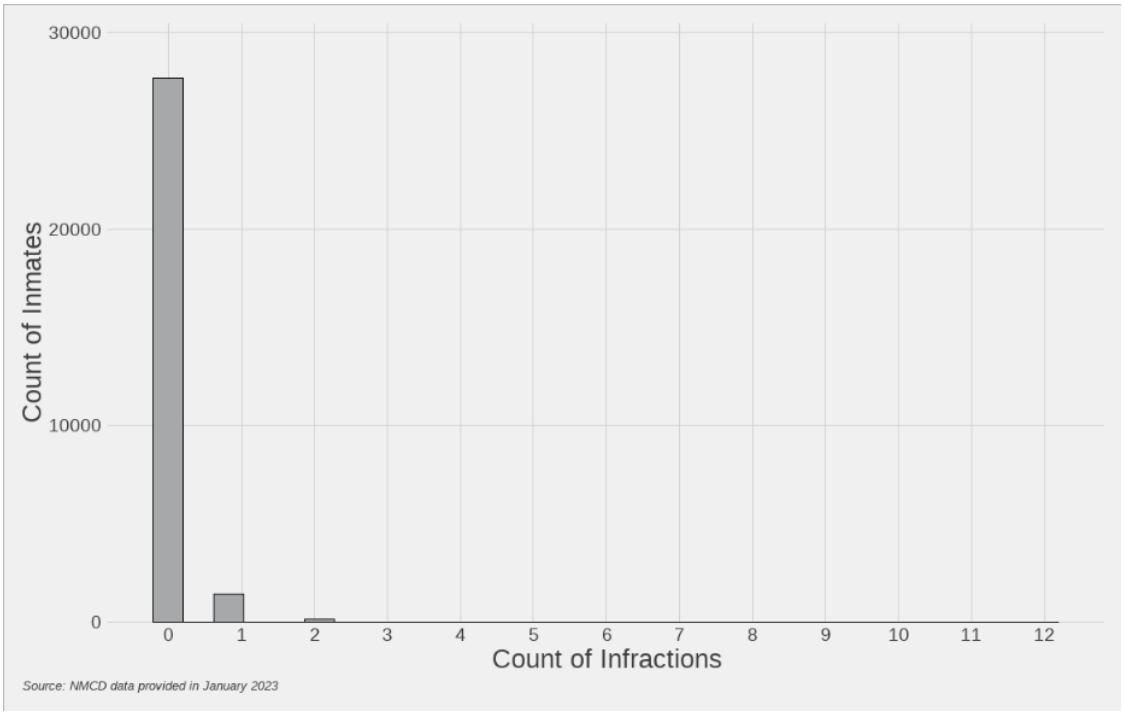


Figure 6.

Count of Inmates by Frequency of SVM Events in Six Months Following Reclassification (n = 29,331)



Figures 3-6 indicate that a small group of inmates committed a disproportionate number of infractions, while most inmates did not engage in misconduct within six months of classification. Among those who did engage in misconduct, most only engaged in it once during the six-months following classification.

We used negative binomial count models to investigate the frequency of GM and SVM and determine whether an inmate's scored custody level, score on the tool, and other control variables – inclusive of race, sex, and programming participation – predicted misconduct. These models adjusted for the inflated zero-count in the dependent variable as most inmates did not engage in misconduct. Our results show that higher tool-generated custody levels and scores predicted increased counts of GM and SVM. Male inmates engaged in significantly higher levels of GM and SVM than female inmates. Additionally, Hispanic inmates engaged in more frequent GM than Whites, Blacks, and Native Americans. However, we found no significant differences in SVM counts across race-ethnicity categories. Participation in prison programming was a significant predictor of overall counts of misconduct within the six-month window following initial classification. We converted the coefficients from the negative binomial count models into predicted misconduct counts by level and present the results in Table 23.

Table 23.
Predicted Counts of GM and SVM by Tool-Generated Level (Classification: Typical Inmate)

	GM	**SVM**
Level I	1.0	0.0
Level II	1.2	0.1
Level III	1.5	0.1
Level IV	1.8	0.1

A Review of NMCD Predictive Validity Statistics – Initial Tool

In Table 24, we present predictive validity statistics used to assess the predictive validity of the initial classification tool stratified by sex and misconduct type including the area under the curve (AUC) statistic derived from a general logistic model predicting misconduct as a function of the score on the tool *across* levels, holding enrollment in programming constant and the 95% confidence interval of the AUC estimate derived from 10,000 bootstrapped samples. In Tables 25 and 26, we report AUCs derived from a general logistic model predicting misconduct as a function of the score on the tool *within* levels, holding enrollment in programming constant and the 95% confidence interval of the AUC estimate derived from 10,000 bootstrapped samples.

Table 24.
Predictive Validity Statistics for the NMCD's Initial Classification Tool

	AUC	**AUC (95% CI)**
Male – GM	0.58	0.56 – 0.60
Male – SVM	0.59	0.55 – 0.63
Female – GM	0.60	0.56 – 0.64
Female – SVM	0.65	0.53 – 0.77

Table 25.
AUCs within Levels for Male Inmates (Initial Classification Tool)

	AUC	**AUC (95% CI)**
Level I (Males – GM)	0.61	0.55 – 0.68
Level II (Males – GM)	0.55	0.52 – 0.56
Level III (Males – GM)	0.54	0.50 – 0.57
Level IV (Males – GM)	0.52	0.40 – 0.63
Level I (Males – SVM)	0.66	0.48 – 0.83
Level II (Males – SVM)	0.56	0.52 – 0.60
Level III (Males – SVM)	0.53	0.45 – 0.60
Level IV (Males – SVM)	0.58	0.44 – 0.72

Table 26.
AUCs within Levels for Female Inmates (Initial Classification Tool)

	AUC	**AUC (95% CI)**
Level I (Females – GM)	0.59	0.53 – 0.65
Level II (Females – GM)	0.58	0.51 – 0.65
Level III (Females – GM)	0.61	0.34 – 0.87
Level IV (Females – GM)	N/A	N/A
Level I (Females – SVM)	0.60	0.41 – 0.77
Level II (Females – SVM)	0.65	0.50 – 0.81
Level III (Females – SVM)	0.60	0.40 – 0.80
Level IV (Females – SVM)	N/A	N/A

The AUC estimates and their bootstrapped confidence intervals suggest that the initial classification tool has poor to fair predictive performance predicting both GM and SVM for both males and females, though considerably more uncertainty exists surrounding the predictive validity of the tool within levels (versus between levels), as given by the width of the bootstrapped confidence intervals. AUC estimates for female inmates in Level IV were inestimable since now females were placed within Level IV custody levels. Generally, the tool had better predictive performance within less restrictive custody levels (e.g., Level I and Level II). The use of overrides by the NMCD limited the statistical power of the present analysis.

Evaluating the Unique Predictive Power of Each Factor – Initial Tool

It is important to see whether the predictive performance of the tool can be improved by better understanding the degree to which tool factors correlate with misconduct. However, guidance on which factors should be included in criminological risk assessments is limited. Speaking to this point, Duwe (2019) states, “The literature offers little, if any, explicit guidance when it comes to revising a risk assessment. The only criterion, albeit an implicit one, seems to be that the challenger—the new, updated assessment—should be able to outperform the incumbent—the older, existing version of the risk assessment instrument” (Duwe 2019: 548). It is important to identify which factors should be included in the classification tool and which should be reweighted or removed to avoid negative statistical consequences. Factors that should be considered for inclusion in a tool are those that are theoretically related to misconduct and which independently predict misconduct.

We used a multifaceted approach to evaluate the predictive performance of factors embedded in the NMCD tools, including bivariate correlations, negative binomial count models, logistic models, and random forest models. In Table 27 and 28, we identify bivariate correlations between score factors and misconduct. In Table 29 and 30, we highlight statistically-significant results of the logistic and negative binomial count models for both male and female inmates. In Appendix C, we present the results of random forest models which quantify the contribution of each tool factor to the tools' overall predictive performance through variable importance plots. In Appendix D, we present predicted probability results from the logistic models highlighted in Tables 30 and 31 which shows how specific factor items predict SVM by sex.

Table 27.

Correlations Between Factors and Counts of Misconduct for Initial Classification Tool (Male Inmates)

	GM	**SVM**
Q1 – Institutional Violence	0.09	0.06
Q2 – Current Conviction Severity	0.00	0.01
Q3 – Escape History	0.02	-0.01
Q4 – Prior Felony Conviction	-0.08	-0.01
Q5 – Prior Conviction Severity	-0.05	0.00
Q6 – Alcohol/Drug Abuse	-0.04	-0.01
Q7 – Current Age	0.25	0.07
Q8 – Gang Membership	0.16	0.07

Table 28.

Correlations Between Factors and Counts of Misconduct Initial Classification Tool (Female Inmates)

	GM	**SVM**
Q1 – Institutional Violence	0.09	0.12
Q2 – Current Conviction Severity	0.11	0.03
Q3 – Escape History	0.0	0.0
Q4 – Prior Felony Conviction	-0.02	-0.01
Q5 – Prior Conviction Severity	0.01	0.04
Q6 – Alcohol/Drug Abuse	0.01	-0.03
Q7 – Current Age	0.26	0.09
Q8 – Gang Membership	0.10	0.13

From Table 27, we observe that for male inmates, the initial classification tool items that correlated the most with GM and SVM were the current age score, gang membership score, and history of institutional violence scores. Similarly, Table 28 shows that for female inmates, the items on the initial classification tool that correlated the most with both GM and SVM were similarly gang membership score, the history of institutional violence score, and current age.

We report on bivariate correlations between factors and misconduct because correlations have historically been reported as metric for evaluating predictive validity in the risk assessment literature. However, it is

important to keep in mind that meaningful predictive inferences based on bivariate correlations are very limited as bivariate correlations are sensitive to scale, the nuances in the ways factors are measured (i.e., the number of response options within a factor can influence how large a correlation appears), and spuriousness which can result in misleading conclusions about the true relationship between factors and misconduct. For these reasons, we present the results of a series of logistic and negative binomial count models in Tables 30 and 31 which statistically control for other factors.

Table 29.

Which Factor Scores Predicted the Presence of Misconduct? (Logistic Results)[5]

	GM – Males	**SVM – Males**	**GM – Females**	**SVM – Females**
Q1 – Institutional Violence	X	X	X	X
Q2 – Current Conviction Severity	--	--	--	--
Q3 – Escape History	--	--	--	--
Q4 – Prior Felony Conviction	--	--	--	--
Q5 – Prior Conviction Severity	X	X	--	--
Q6 – Alcohol/Drug Abuse	--	--	--	--
Q7 – Current Age	X	X	X	X
Q8 – Gang Membership	X	X	--	X

Table 30.

Which Factor Scores Predicted Higher Counts of Misconduct? (Negative Binomial Results)

	GM – Males	**SVM – Males**	**GM – Females**	**SVM – Females**
Q1 – Institutional Violence	X	X	X	X
Q2 – Current Conviction Severity	--	--	X	--
Q3 – Escape History	--	--	--	--
Q4 – Prior Felony Conviction	--	--	--	--
Q5 – Prior Conviction Severity	--	--	--	--
Q6 – Alcohol/Drug Abuse	X	--	--	--
Q7 – Current Age	X	X	X	X
Q8 – Gang Membership	X	X	X	X

There were sex differences in the factors that predicted misconduct occurrence and count in the six months following initial classification. Specifically, fewer factors were significant predictors for female SVM occurrence and count compared to male SVM. Only institutional violence, current age, and gang membership scores were independently predictive of SVM among female inmates. Additionally, prior conviction severity was a significant predictor of male GM and SVM occurrence, but not for female GM or SVM occurrence. However, prior conviction severity significantly predicted a higher GM count for female inmates. Across both male and female inmates, factors such as escape history, alcohol and drug

[5] Reference for comparison is lowest point value of the specific variable (e.g., 0 for No Gang Membership for Q8).

abuse, and prior felony convictions were not unique predictors of either SVM or GM occurrence. In contrast, current age, gang membership, and history of institutional violence emerged as particularly predictive of GM or SVM occurrence.

We present a detailed analysis of how different factor response levels predicted SVM in Appendix D which provides important nuance for better understanding why only examining bivariate correlations or statistical significance is, in isolation, an insufficient approach. The revisions we propose to the scoring of the existing tools reflect inexact considerations which aim to balance (1) the predicted probabilities of SVM associated with each factor-level given in Appendix D against (2) the overall statistical significance of the factor and (3) practical institutional and policy-related considerations (e.g., whether a factor should be deleted due inadequate face validity, as is the case with the alcohol/drug abuse factor).

In creating revisions to the scoring of the revised tools, we assigned higher point totals to factors which more strongly predicted SVM and which were statistically significant predictors and balanced this with broader institutional goals (e.g., including factors for reputation management reasons, such as escape history, even if their ability to predict misconduct is unknown given the rarity of the event within the sampled period). We want to underscore that only examining statistical significance at the factor-level in prior tables is inadequate for making determinations about the relative scoring of items embedded within factors (e.g., how many points to assign within factor levels). In Appendix D, we provide more context on the different misconduct profiles of male and female inmates including the predicted probabilities of engaging in SVM.

A Review of NMCD Predictive Validity Statistics – Reclassification Tool

In Table 31, we present predictive validity statistics used to assess the predictive validity of the reclassification tool stratified by sex and misconduct type including the AUC statistic derived from a general logistic model predicting misconduct as a function of the score on the tool across levels, holding enrollment in programming constant and the 95% confidence interval of the AUC estimate derived from 10,000 bootstrapped samples. In Tables 32 and 33, we report the same but the AUCs within levels.

Table 31.

Predictive Validity Statistics for the NMCD's Reclassification Tool

	AUC	**AUC (95% CI)**
Male – GM	0.56	0.55 – 0.56
Male – SVM	0.60	0.58 – 0.61
Female – GM	0.70	0.68 – 0.72
Female – SVM	0.71	0.66 – 0.76

Table 32.
AUCs within Levels for Male Inmates (Reclassification Tool)

	AUC	AUC (95% CI)
Level I (Males – GM)	0.61	0.59 – 0.64
Level II (Males – GM)	0.57	0.56 – 0.59
Level III (Males – GM)	0.52	0.51 – 0.53
Level IV (Males – GM)	0.52	0.51 – 0.54
Level I (Males – SVM)	0.69	0.61 – 0.77
Level II (Males – SVM)	0.57	0.54 – 0.59
Level III (Males – SVM)	0.53	0.51 – 0.55
Level IV (Males – SVM)	0.52	0.49 – 0.55

Table 33.
AUCs within Levels for Female Inmates (Reclassification Tool)

	AUC	AUC (95% CI)
Level I (Females – GM)	0.63	0.58 – 0.69
Level II (Females – GM)	0.59	0.56 – 0.63
Level III (Females – GM)	0.52	0.49 – 0.57
Level IV (Females – GM)	0.52	0.42 – 0.62
Level I (Females – SVM)	0.67	0.43 – 0.83
Level II (Females – SVM)	0.58	0.49 – 0.67
Level III (Females – SVM)	0.62	0.54 – 0.70
Level IV (Females – SVM)	0.57	0.42 – 0.71

Compared to the initial classification tool, the reclassification tool generally was more predictive of misconduct using AUC as the comparison metric, with less variability – on balance – surrounding estimates. The reclassification tool includes factors that assess an inmate's recent behavior in prison (e.g., disciplinary history) which tend to be more predictive of future behavior which may, in part, account for the observed increase in AUCs.

Evaluating the Unique Predictive Power of Each Factor – Reclassification Tool

The reclassification tools' AUCs generally possessed fair to good predictive quality for male inmates , ranging from 0.57 to 0.64 with relatively tight confidence intervals. There was generally more uncertainty bounding AUC estimates for female inmates across levels, reflecting – in part – the smaller sample size of female inmates.

Table 34.

Correlation Between Factors and Counts of Misconduct for Reclassification Tool (Male Inmates)

	GM	**SVM**
Q1 – Institutional Violence	0.00	0.08
Q2 – Risk of Current Conviction	-0.04	0.01
Q3 – Escape History	0.00	0.02
Q4 – Prior Felony Convictions	-0.09	-0.00
Q5 – Disciplinary Convictions Received	0.05	0.04
Q6 – Clear Conduct	0.10	0.04
Q7 – Program/Work Performance	0.07	0.04
Q8 – Current Age	0.20	0.06
Q9 – Gang Membership	0.05	0.03

Table 35.

Correlation Between Factors and Counts of Misconduct for Reclassification Tool (Female Inmates)

	GM	**SVM**
Q1 – Institutional Violence	0.20	0.12
Q2 – Risk of Current Conviction	0.05	0.05
Q3 – Escape History	0.00	0.02
Q4 – Prior Felony Convictions	-0.06	0.00
Q5 – Disciplinary Convictions Received	0.26	0.12
Q6 – Clear Conduct	0.14	0.05
Q7 – Program/Work Performance	0.09	0.00
Q8 – Current Age	0.17	0.05
Q9 – Gang Membership	0.10	0.02

From Tables 34 and 35, we observe that the strongest correlates of SVM were history of institutional violence and age for both male and female inmates. These emerged as significant factors for both male and female inmates. We note that a larger sample size provides more data points and reduces the impact of random variation in the data. This means that with a larger sample size, statistical hypothesis tests are better able to distinguish between a true effect and random noise. As a result, the statistical power of the test increases which makes it easier to detect a statistically significant effect. However, this decreases the usefulness of using statistical significance as a method for understanding which predictor variables should be included in proposed revisions to the existing tool. As with the initial classification tool, the number of significant predictors was smaller for female misconduct following reclassification than for male misconduct.

Tables 36 and 37 show the logistic and negative binomial count model results for both male and female inmates for the reclassification tool.

Table 36.
Which Factors Predicted the Presence of Misconduct? (Logistic Results)[6]

	GM – Males	**SVM – Males**	**GM – Females**	**SVM – Females**
Q1 – Institutional Violence	X	X	X	--
Q2 – Risk of Current Conviction	X	--	X	--
Q3 – Escape History	--	--	--	--
Q4 – Prior Felony Conviction	--	--	--	--
Q5 – Disciplinary Convictions Received	X	--	X	X
Q6 – Clear Conduct	X	--	--	--
Q7 – Program/Work Performance	X	X	X	--
Q8 – Current Age	X	X	X	X
Q9 – Gang Membership	X	--	X	--

Table 37.
Which Factors Predicted Higher Counts of Misconduct? (Negative Binomial Results)

	GM – Males	**SVM – Males**	**GM – Females**	**SVM – Females**
Q1 – Institutional Violence	X	X	X	--
Q2 – Risk of Current Conviction	X	--	X	--
Q3 – Escape History	--	X	--	--
Q4 – Prior Felony Conviction	--	X	X	--
Q5 – Disciplinary Convictions Received	X	X	X	X
Q6 – Clear Conduct	--	--	--	--
Q7 – Program/Work Performance	X	X	X	--
Q8 – Current Age	X	X	X	X
Q9 – Gang Membership	X	--	X	--

As with the initial classification tool, from Tables 36 and 37 we observe that there were differences in the number of predictors of male and female SVM: specifically, female inmates' SVM was predicted by fewer factors than male SVM, consistent with the existing evidence-base which suggests that female SVM is rarer and predicted by fewer factors. Escape history, clear conduct, program and work performance, and gang membership did not significantly predict SVM within female inmates following reclassification whereas these factors significantly predicted SVM within male inmates following reclassification.

[6] Reference for comparison is lowest point value of the specific variable (e.g., 0 for No Gang Membership for Q9).

Knowing that a factor was a statistically significant predictor of misconduct – while important – does not provide insight into how the different scored response options within a factor's measurement related to misconduct. As one example, consider the factor history of institutional violence on the reclassification tool for male inmates. This variable is scored as follows on the currently used reclassification tool:

- None = 0 points
- Nonviolent/Serious Class A Level Incidents = 1 point
- Violent Incident with No Weapon = 3 points
  Violent Incident with Weapon = 6 points

On the reclassification tool, inmates who had the highest risk of current conviction score were assigned the same number of points, 6, as inmates with a history of institutional violence with a weapon, even though inmates with a history of institutional violence had higher predicted probabilities of engaging in SVM than inmates who had the highest risk of current conviction [8.2% SVM risk for inmates with history of institutional violence with a weapon versus 4% SVM risk for highest current conviction severity]. Given that the factors are scored similarly yet have substantively different relationships to both GM and SVM risk and given the higher VIP scores of the history of institutional violence factor relative to current conviction severity, we recommend increasing point totals for the institutional violence history factor in relation to the current conviction severity factor. Results such as this raise important considerations which motivate our proposed revisions of the existing tools. We return to this conversation in more detail in the *Recommendations* section and Appendix D.

Overrides and Misconduct

52% of initial classification records (*n* = 5,767) and 42% of reclassification records (*n* = 21,626) were ineligible for inclusion in the validation sample due to the use of overrides. In the admission cohort, 93% of overrides were mandatory overrides with a majority being cases where individuals who scored at lower levels were placed at a higher-custody level than their scoring on the tool would recommend due to "Mental Health/Medical" restrictions, mostly occurring among the subset of individuals who scored at Level I who were subsequently overridden typically into Level II facilities (29%; *n* = 7,901). Table 40 displays the most common mandatory override factor at each level, and Table 41 displays the custody levels overridden inmates were moved to.

Table 40.
Most Common Mandatory Overrides by Level

	Most Common Reason for Mandatory Override
Level I	Mental Health/Medical: 28% (*n* = 4,221)
Level II	Felony Detainer: 41% (*n* = 2,910)
Level III	Validated or Suspected STG Affiliation: 8% (*n* = 153)

Table 41.

What Custody Levels Were Overridden Inmates Placed At?

	Level I	**Level II**	**Level III**	**Level IV**
Level I (Tool Level)	--	65% (*n* = 9,843)	34% (*n* = 5,211)	1% (*n* = 201)
Level II (Tool Level)	1% (*n* = 02)	--	90% (*n* = 6,394)	8% (*n* = 586)
Level III (Tool Level)	0% (*n* = 0)	56% (*n* = 1,018)	--	44% (*n* = 809)
Level IV (Tool Level)	0% (*n* = 0)	0% (*n* = 0)	100% (*n* = 3,230)	--

Logistic regression indicated that female, older, and non-Hispanic White inmates were more likely to be overridden into higher custody levels in the initial classification sample. Similar results were found in the reclassification sample, with the addition of Native American inmates being more likely to be overridden.

To evaluate the relationship between the use of overrides and misconduct, we compared misconduct counts of inmates who were not overridden to inmates who were. Table 42 summarizes misconduct counts within and across custody levels across all inmate classification records. Rows identify the level an inmate scored on the tool and columns indicate the level the inmate was placed at. The red diagonal indicates the misconduct counts of inmates who were placed in custody levels consistent with their tool scores.

Table 42.

Average GM Count by Level (All Inmates)

	Level I (Placed)	**Level II (Placed)**	**Level III (Placed)**	**Level IV (Placed)**
Level I (Tool)	0.94	0.56	0.54	0.15
Level II (Tool)	1.17	2.41	1.10	0.39
Level III (Tool)	--	1.48	2.87	0.72
Level IV (Tool)	--	--	2.10	2.20

Table 42 demonstrates that inmates who were overridden into higher custody levels than their tool-generated levels engaged in less misconduct than those who scored at the same level they were overridden into. These findings reveal that inmates who were mandatorily overridden tended to have misconduct profiles more like inmates at their scored level rather than inmates at the level they were overridden into. The disparity in misconduct profiles is similar when we consider cases where inmates were overridden into more restrictive custody levels due to medical and mental health restrictions. We present misconduct counts in Table 43.

Table 43.
Misconduct Profiles of Inmates Mandatorily Overridden Because of Medical/Mental Health Restrictions

	Misconduct (Not Overridden)	**Misconduct (Overridden)**
Level I (Tool)	0.94	0.56
Level II (Tool)	2.41	2.27

Table 43 shows that inmates who were overridden from lower custody levels because of medical and mental health restrictions (i.e., Level I into Level II or Level III; Level II into Level III) tended to have misconduct profiles like inmates at the level they scored at on the tool and, in the case of Level II inmates, even lower counts of misconduct. We discuss implications of this specific policy in the *Recommendations* subsection.

Data Analysis – Proposed Scoring Revisions to the Classification Tools

Literature on the psychometrics of scale construction (Boateng et al., 2018) and criminological risk assessment (Coid et al., 2011; Chen at al., 2019) note it is important to include variables predictive of misconduct in risk forecasting tools. We need to develop scores for factors included on the tool which accurately reflect the extent to which the factor predicted misconduct risk. To this end, we used predicted probabilities generated from a logistic regression predicting SVM as a function of each tool factor (as previewed in Tables 32 and 40), rank-ordered predicted probabilities across all factor response options and generated new cumulative scores for the revised assessment. We then customized these scores based on a few practical considerations and defined level thresholds. We present a new version of the tool in Appendix F.

In revising factor scores, we assume that the reduction of SVM – versus GM – is the main objective of the classification tool. We assume this because literature on institutional misconduct suggests SVM is more costly and disruptive to prison operations than GM (Wooldredge 2020). We discussed this point with NMCD classification staff in April 2023. We also note that the small number of inmates who received points for some factor response options made it difficult to evaluate the relationship between statistically-rare factors and misconduct (e.g., escape attempts are rare). For example, only 2% of male inmates at initial classification and less than 1% of female inmates at initial classification had any previous escape history. With few instances of a rare event like escape, there is insufficient statistical power to detect meaningful associations or to draw reliable conclusions about the relationship between the factor, such as escape history, and SVM.

Recommendations

We recognize the NMCD faces complex, competing interests determining how to classify inmates. There are many considerations, such as the need to balance predictive accuracy with organizational constraints like facility capacity, services, and staffing. Additionally, tolerance for different classification errors is subjective and varies based on an organization's risk aversion.

In general, the NMCD incorporates essential components of an objective prison classification system and specifically, the reclassification tool has “good” predictive performance. We recommend the NMCD

adopt the following recommendations to the extent possible within a reasonable time frame given internal technological constraints which may limit the adoption speed of proposed recommendations.

1. *Remove the Use of Mandatory Overrides to Acceptable Levels*: We suggest removing medical and mental health overrides and felony detainer mandatory overrides. Currently, over half of initial classification records and almost half of reclassification records were overridden through mandatory overrides. The frequent use of overrides restricted our ability to evaluate the predictive validity of the initial classification tool. Furthermore, misconduct counts were lower for inmates who were overridden into higher custody levels compared to those who scored at the level they were overridden to. The ability to use a validated classification tool reliably is limited by high override rates. We understand it may take time to reduce the use of mandatory overrides for several reasons including staffing limitations, the availability of mental health services and appropriate programming for inmates in all facilities and security levels, facility design capacities, and the need to adapt policy and practice. Where practicable, we advise transferring inmates across facilities *within* the same custody level to provided needed services or imposing necessary restrictions *within* levels instead of overriding inmates into higher custody-level facilities. If the NMCD continues to override inmates into higher custody levels just because the services inmates need are unavailable at the level they are classified into, this threatens to violate the predictive accuracy of the tools and fails to address inmates' needs in accordance with a risk-needs-response approach.
2. *Continue Evaluating Tool Implementation and Validity*: Classification tools must continually be evaluated to ensure they remain effective and fair over time. This involves monitoring the tools' performance and making updates as needed to ensure tools accurately predict an inmate's risk of misconduct. We recommend formally revalidating the classification tools every five years. We note that the scoring revisions and level thresholds we propose do not take into consideration facility bedding availability and indeed, cannot be retrospectively evaluated due to a lack of direct access to CMIS data. It will be necessary to understand whether and how the revised classification tool shifts the distribution of inmates across custody levels and facilities by prospectively evaluating the tool on new samples alongside the current tool and making revisions to the level thresholds if needed. Further, if multiple report recommendations are taken, it is possible that misconduct outcomes at the facility-level may change in unpredictable ways (e.g., a reduction in mandatory overrides may impact misconduct levels at different levels in ways that could not be anticipated by the proposed revisions to the tools). Additionally, it will be important to assess predictive performance by race-ethnicity to evaluate predictive parity. We propose continual monitoring of the new tool while it is implemented to evaluate bedding and facility needs and descriptive features of the tool in real-time.
3. *Change Existing Factor Scores*: We recommend revising the custody scoring forms by attaching greater weight to factors that are more predictive of SVM risk and reducing scores of less predictive factors or removing non-predictive factors. We provide specific details on recommended scoring changes in Appendix F.
4. *Reduce Time Frames for Some Factors*: We recommend reducing the timeframes used for factors that evaluate an inmate's past behavior, such as the history of institutional violence factor, which currently reviews an individual's misconduct background for 10 years prior to classification. We suggest reducing this and other time frames to three years prior to the classification event since

more recent behavior is a better predictor of future behavior than distant past behavior. We provide a more in-depth explanation – both theoretically and empirically - for why we propose shortening lookback timeframes for different factors in Appendix E.

5. *Remove the Alcohol/Drug Abuse Factor*: The alcohol/drug abuse factor on the initial classification tool does not measure alcohol or drug abuse; it measures drug trafficking and distribution. Per bivariate and multivariate analyses, this factor was not predictive of GM or SVM. We recommend removing the current alcohol/drug abuse measure on the initial classification tool.
6. *Remove the History of Disciplinary Factor*: The history of disciplinary factor on the reclassification tool was not a unique predictor of GM or SVM, and scores for this factor were non-linearly related to misconduct (see Appendix F). Further, this factor was correlated with and indirectly measured by the history of institutional violence factor and disciplinary convictions received factor. It is generally advised to not include redundant items in scale construction. We recommend removing the history of disciplinary factor on the reclassification tool.
7. *Remove the Program/Work Performance Factor or Appropriately Measure*: The measurement of program/work performance reflecting maximum good time achieved is not an appropriate way to evaluate program and work performance and can exclude inmates from receiving tool point deductions conditional on their facility placement (i.e., the capacity to participate in programs varies by custody level). Additionally, this factor was not predictive of GM or SVM. We recommend removing this factor from the reclassification tool.

Conclusions

The focus groups and observations highlight that the NMCD has the essential components of an objective classification system. Our empirical validation found that the reclassification tool was more predictive of misconduct than the initial classification tool for both male and female inmates and generally had good classification performance, though there are limitations associated with using the AUC as a primary metric for evaluating predictive validity. Due to the small sample size of female inmates at higher custody levels, we could not confidently evaluate the predictive validity of the initial tool among female inmates. Because there were factors on the initial and reclassification tools that were more strongly related to SVM, we recommend increasing point totals for these factors, as outlined in Appendix F and as justified in Appendix C, Appendix D, and Appendix E. Our analysis did not reveal significant differences in the proportion of male and female inmates within each level who engaged in SVM, nor did it reveal statistically significant differences in the predicted probabilities of SVM for male and female inmates by factor on both the initial and reclassification tools, suggesting that the use of only one form for both male and female is empirically grounded. The frequent use of mandatory overrides impacts the usefulness of the classification and reclassification tools by changing the distribution of risk-profiles of inmates within any given custody level. Because of this, we recommend the NMCD offer needed behavioral health and medical services at lower custody levels or, if not practicable, remove the use of mandatory overrides for medical and mental health reasons, allowing the raw score on the tool to determine custody placement. While the existing tools have now been validated, we encourage the NMCD to continually monitor and validate the existing tools and any revisions made to them over the coming years.

References

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, *17*, 19-52. https://doi.org/10.1177/0093854890017001004

Austin, J., & Hardyman, P. L. (2004). *Objective prison classification: A guide for correctional agencies*. US Department of Justice, National Institute of Corrections.

Berk, R. A., Kriegler, B., & Baek, J. H. (2006). Forecasting dangerous inmate misconduct: An application of ensemble statistical procedures. *Journal of Quantitative Criminology*, *22*(2), 131-145.

Berk, R. A., & Bleich, J. (2013). Statistical procedures for forecasting criminal behavior: A comparative assessment. *Criminology & Public Policy*, 12, 513-544.

Blowers, A. N., & Blevins, K. R. (2015). An examination of prison misconduct among older inmates. *Journal of Crime and Justice*, *38*(1), 96-112.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health*, *6*, 149, 1-18.

Bonner, H. S., Rodriguez, F. A., & Sorensen, J. R. (2017). Race, ethnicity, and prison disciplinary misconduct. *Journal of Ethnicity in Criminal Justice*, *15*(1), 36-51.

Bosma, A. Q., C. van Ginneken, E. F., Sentse, M., & Palmen, H. (2020). Examining inmate misconduct: A multilevel test using personal characteristics, prison climate, and prison environment. *Crime & Delinquency*, *66*(4), 451-484.

Butler, M., McNamee, C. B., & Kelly, D. (2021). Exploring prison misconduct and the factors influencing rule infraction in Northern Ireland. *European Journal on Criminal Policy and Research*, 1-20. https://doi.org/10.1007/s10610-021-09491-6

Butler, H. D., Caudill, J. W., Craig, J. M., DeLisi, M., & Trulson, C. R. (2021). 99 percenters: An examination of the misconduct careers of the most violent and disruptive incarcerated delinquents. *Aggression and Violent Behavior*, *60*, 101520.

Celinska, K., & Sung, H. E. (2014). Gender differences in the determinants of prison rule violations. *The Prison Journal*, *94*(2), 220-241.

Chen, J. C., Dunn, A., Hood, K., Driessen, A., & Batch, A. (2019). Off to the races: A comparison of machine learning and alternative data for predicting economic indicators. In *Big Data for 21st Century Economic Statistics*. University of Chicago Press.

Cochran, J. C., Mears, D. P., Bales, W. D., & Stewart, E. A. (2014). Does inmate behavior affect post-release offending? Investigating the misconduct-recidivism relationship among youth and adults. *Justice Quarterly*, *31*(6), 1044-1073.

Cochran, J. C., & Mears, D. P. (2017). The path of least desistance: Inmate compliance and recidivism. *Justice Quarterly*, *34*(3), 431-458.

Coid, J. W., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Farrington, D., & Rogers, R. (2011). Most items in structured risk assessment instruments do not predict violence. *The Journal of Forensic Psychiatry & Psychology*, *22*(1), 3-21.

DeMaris, A., & Selman, S. H. (2013). Logistic regression. In *Converting Data into Evidence* (pp. 115-136). Springer, New York, NY.

Desmarais, S. L., & Singh, J. P. (2013). *Risk assessment instruments validated and implemented in correctional settings in the United States* (Tech. Rep.). Lexington, KY: Council of State Governments Justice Center. Retrieved from https://csgjusticecenter.org/ wp-content/uploads/2014/07/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf

Desmarais, S. L., Zottola, S. A., Duhart Clarke, S. E., & Lowder, E. M. (2021). Predictive validity of pretrial risk assessments: A systematic review of the literature. *Criminal Justice and Behavior*, 48(4), 398-420.

Diamond, B., Morris, R. G., & Barnes, J. C. (2012). Individual and group IQ predict inmate violence. *Intelligence*, *40*(2), 115-122.

Drury, A. J., & DeLisi, M. (2010). The past is prologue: Prior adjustment to prison and institutional misconduct. *The Prison Journal*, *90*(3), 331-352.

Duwe, G., & Rocque, M. (2017). The effects of automating recidivism risk assessment on reliability, predictive validity, and return on investment (ROI). *Criminology & Public Policy*, *16*, 235–269.

Duwe, G., & Kim, K. (2016). Sacrificing accuracy for transparency in recidivism risk assessment: The impact of classification method on predictive performance. *Corrections*, *1*(3), 155-176.

Fazel, S., Burghart, M., Fanshawe, T., Gil, S. D., Monahan, J., & Yu, R. (2022). The predictive performance of criminal risk assessment tools used at sentencing: Systematic review of validation studies. *Journal of Criminal Justice*, 81, 101902.

Georgiou, G. (2019). Weights matter: Improving the predictive validity of risk assessments for criminal offenders. *Journal of Offender Rehabilitation*, *58*(2), 92-116.

Gottfredson, S. D., & Moriarty, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency*, *52*(1), 178-200.

Hamilton, Z., Kigerl, A., Campagna, M., Barnoski, R., Lee, S., Van Wormer, J., & Block, L. (2016). Designed to fit: The development and validation of the STRONG-R recidivism risk assessment. *Criminal Justice and Behavior*, *43*(2), 230-263.

Hamilton, Z., Neuilly, M. A., Lee, S., & Barnoski, R. (2015). Isolating modeling effects in offender risk assessment. *Journal of Experimental Criminology*, *11*(2), 299-318.

Harer, M. D., & Langan, N. P. (2001). Gender differences in predictors of prison violence: Assessing the predictive validity of a risk classification system. *Crime & Delinquency*, *47*(4), 513-536.

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. *Psychological Assessment*, *21*(1), 1-21.

Ho, T., & Rocheleau, G. C. (2021). Gender differences in risk assessment among youth adjudicated to juvenile correctional facilities. *Women & Criminal Justice*, *31*(6), 463-475.

Kwegyir-Aggrey, K., Gerchick, M., Mohan, M., Horowitz, A., & Venkatasubramanian, S. (2023). The Misuse of AUC: What High Impact Risk Assessment Gets Wrong. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1570-1583).

Labrecque, R. M., Mears, D. P., & Smith, P. (2020). Gender and the effect of disciplinary segregation on prison misconduct. *Criminal Justice Policy Review*, *31*(8), 1193-1216.

Labrecque, R. M., Smith, P., Lovins, B. K., & Latessa, E. J. (2014). The importance of reassessment: How changes in the LSI-R risk score can improve the prediction of recidivism. *Journal of Offender Rehabilitation*, *53*(2), 116-128.

Logan, M. W., Long, J., DeLisi, M., & Hazelwood, A. R. (2023). Serious, Violent, and Chronic Prison Misconduct: Are the Predictors the Same for Women and Men?. *The Prison Journal*, 103(1), 23-44.

Long, J. S. (2020). Appropriate classification of inmates: Balancing prison safety with the least restrictive placements of Ohio inmates (Doctoral dissertation, University of Cincinnati).

Lowenkamp, C. T., & Bechtel, K. (2007). The predictive validity of the LSI-R on a sample of offenders drawn from the records of the Iowa Department of Correction Data Management System. *Fed. Probation*, 71, 25.

Lovins, B. K., Latessa, E. J., May, T., & Lux, J. (2018). Validating the Ohio Risk Assessment System community supervision risk assessment instrument with a diverse sample from Texas. *Corrections*, *3*(3), 186-202.

Matthews, B., & Hubbard, D. J. (2008). Moving ahead: Five essential elements for working effectively with girls. *Journal of Criminal Justice*, *36*, 494-502.

Morgan, A. (2018). How Much Does Prison Really Cost?: Comparing the Costs of Imprisonment with Community Corrections.

Papalia, N., Spivak, B., Daffern, M., & Ogloff, J. R. (2019). A meta-analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clinical Psychology: Science and Practice*, *26*(2), e12282.

Pompoco, A., Wooldredge, J., Lugo, M., Sullivan, C., & Latessa, E. J. (2017). Reducing inmate misconduct and prison returns with facility education programs. *Criminology & Public Policy*, *16*(2), 515-547.

Pyrooz, D. C., Turanovic, J. J., Decker, S. H., & Wu, J. (2016). Taking stock of the relationship between gang membership and offending: A meta-analysis. *Criminal Justice and Behavior*, *43*(3), 365-397.

Rocheleau, A. M. (2014). Inmates' coping skills and involvement in serious prison misconduct. *Victims & Offenders*, *9*(2), 149-177.

Semenza, D. C., & Grosholz, J. M. (2019). Mental and physical health in prison: how co-occurring conditions influence inmate misconduct. *Health & Justice*, *7*(1), 1-12.

Severson, R. E. (2019). Gender differences in mental health, institutional misconduct, and disciplinary segregation. *Criminal Justice and Behavior*, *46*(12), 1719-1737.

Singh, J. P., Desmarais, S. L., & Van Dorn, R. A. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences & the Law*, *31*(1), 55-73.

Sorensen, J. R., & Cunningham, M. D. (2007). Operationalizing risk: The influence of measurement choice on the prevalence and correlates of prison violence among incarcerated murderers. *Journal of Criminal Justice*, *35*(5), 546-555.

Steiner, B., & Cain, C. M. (2016). The relationship between inmate misconduct, institutional violence, and administrative segregation: A systematic review of the evidence. *Restrictive housing in the US: Issues, challenges, and future directions*, 165-197.

Steiner, B., Ellison, J. M., Butler, H. D., & Cain, C. M. (2017). The impact of inmate and prison characteristics on inmate victimization. *Trauma, Violence, & Abuse*, *18*(1), 17-36.

Steiner, B., & Wooldredge, J. (2014). Sex differences in the predictors of inmate misconduct. *Criminal Justice and Behavior*, *41*(4), 433-452.

Steiner, B., Butler, H. D., & Ellison, J. M. (2014). Causes and correlates of prison inmate misconduct: A systematic review of the evidence. *Journal of Criminal Justice*, *42*(6), 462-470.

Tahamont, S. (2019). The effect of facility security classification on serious rules violation reports in California prisons: a regression discontinuity design. *Journal of Quantitative Criminology*, *35*(4), 767-796.

Tewksbury, R., Connor, D. P., & Denney, A. S. (2014). Disciplinary infractions behind bars: An exploration of importation and deprivation theories. *Criminal Justice Review*, *39*, 201–218

Thomson, N. D., Vassileva, J., Kiehl, K. A., Reidy, D., Aboutanos, M., McDougle, R., & DeLisi, M. (2019). Which features of psychopathy and impulsivity matter most for prison violence? New evidence among female inmates. *International Journal of Law and Psychiatry*, *64*, 26-33.

Thomson, N. D., Towl, G. J., & Centifanti, L. (2016). The habitual female offender inside: How psychopathic traits predict chronic prison violence. *Law and Human Behavior*, *40*(3), 257-269.

Ting, M. H., Chu, C. M., Zeng, G., Li, D., & Chng, G. S. (2018). Predicting recidivism among youth offenders: Augmenting professional judgement with machine learning algorithms. *Journal of Social Work*, *18*(6), 631-649.

Valentine, C. L., Mears, D. P., & Bales, W. D. (2015). Unpacking the relationship between age and prison misconduct. *Journal of Criminal Justice*, *43*(5), 418-427.

Walters, G. D., & Crawford, G. (2013). In and out of prison: Do importation factors predict all forms of misconduct or just the more serious ones?. *Journal of Criminal Justice*, *41*(6), 407-413.

Walters, G. D., & Crawford, G. (2014). Major mental illness and violence history as predictors of institutional misconduct and recidivism: Main and interaction effects. *Law and Human Behavior*, *38*(3), 238-247.

Wang, C., Han, B., Patel, B., & Rudin, C. (2022). In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction. *Journal of Quantitative Criminology*, 1-63.

Wooldredge, J., & Steiner, B. (2013). Violent victimization among state prison inmates. *Violence and Victims*, *28*(3), 531-551.

Wooldredge, J. (2020). Prison culture, management, and in-prison violence. *Annual Review of Criminology*, 3, 165-188.

Worrall, J. L., & Morris, R. G. (2011). Inmate custody levels and prison rule violations. *The Prison Journal*, *91*(2), 131-157.

Wright, E. M., Van Voorhis, P., Salisbury, E. J., & Bauman, A. (2012). Gender-responsive lessons learned and policy implications for women in prison: A review. *Criminal Justice and Behavior*, 39(12), 1612-1632.

Appendix A – Current NMCD Classification Tools

Form CD-081200.1
Reviewed/Revised 03/31/22

NEW MEXICO CORRECTIONS DEPARTMENT
INITIAL CUSTODY SCORING FORM

Inmate's Name:_____ NMCD#_____
Last *First* *MI*

Classification Officer:_____ Classification Date:_____

1. **HISTORY OF INSTITUTIONAL ADJUSTMENT/VIOLENCE.** (Review individual's entire background for 10 years prior to classification date to include juvenile incidents) (Include date of incident; rate most severe)
   None _____ 0
   Ten or more non-violent disciplinary reports _____ 2
   Non-Violent /Serious Class A Level Incidents _____ 2
   Violent Incident with no weapon, serious injury or death _____ 4
   Violent Incident involving a weapon, serious injury or death _____ 7

2. **RISK OF CURRENT CONVICTION** (score the most serious conviction, list offense and date)
   Low _____ 1
   Moderate _____ 3
   High _____ 5
   Highest _____ 7

3. **ESCAPE HISTORY** (Last 10 years from this rating date. List date of escape)
   None _____ 0
   Escape/Attempted Escape from Level I or II, County Jail, Juvenile Facility, or Peace Officer (no violence) _____ 4
   Escape/Attempted Escape from Level III facility or above (no violence) _____ 4
   Escape/Attempted Escape (with violence) _____ 7

   SUB-TOTAL FOR ITEMS 1-3 (If points equal 10 or more, classify as Level IV)

4. **PRIOR FELONY CONVICTIONS** (Do not include current conviction; list offenses and dates.)
   None _____ 0 One ____ 1 Two ____ 2 Three or more ____ 3

5. **RISK OF PRIOR CONVICTIONS** (Score the most serious offence; list offence and date)
   None/Low ___ 0 Moderate _____ 1 High ____ 2 Highest _ 3

6. **ALCOHOL/DRUG ABUSE** (Last 10 years from this rating date; list offense and date)
   None _____ 0
   Trafficking or Distribution of Alcohol/Drugs _____ 1

7. **CURRENT AGE**
   21 and under _ 2 22 to 25 _ 1 26 to 34 _ 0 35 to 44 _ -1 45 and above _____ -2

8. **GANG MEMBERSHIP/ACTIVITIES IN THE PAST 10 YEARS**
   Yes _____ 2 No _____ 0

   TOTAL SCORE (Add 1 through 8)

Form CD-081200.2
Reviewed/Revised 03/31/22

NEW MEXICO CORRECTIONS DEPARTMENT
RECLASSIFICATION SCORING FORM

Inmate's Name:_____NMCD#_____
Last *First* *MI*

Classification Officer:_____Reclassification Date:_____

1. **HISTORY OF INSTITUTIONAL ADJUSTMENT/VIOLENCE.** (Review individual's entire background for 10 years prior to classification date to include juvenile incidents) (Include date of incident; rate most severe)
   None _____ 0
   Non-Violent /Serious Class A Level Incidents _____ 1
   Violent Incident with no weapon, serious injury or death _____ 3
   Violent Incident involving a weapon, serious injury or death _____ 6

2. **RISK OF CURRENT CONVICTION** (score the most serious conviction, list offense and date)
   Low _____ 1
   Moderate _____ 2
   High _____ 4
   Highest _____ 6

3. **ESCAPE HISTORY** (Last 10 years from this rating date. List date of escape)
   None _____ 0
   Escape/Attempted Escape from Level I or II, County Jail, Juvenile Facility, or Peace Officer (no violence) _____ 3
   Escape/Attempted Escape from Level III facility or above (no violence) _____ 5
   Escape/Attempted Escape (with violence) _____ 6

   SUB-TOTAL FOR ITEMS 1-3 (If points equal 10 or more, classify as Level IV)

4. **PRIOR FELONY CONVICTIONS** (Do not include current conviction; list offenses and dates.)
   None _____ 0 One ____ 1 Two ____ 2 Three or more ____ 3

5. **DISCIPLINARY CONVICTIONS RECEIVED** (only one per class)
   Last 24 months: One or more class A = 5 _____
   Last 24 months: One or more class B = 4 _____
   Last 24 months: One or more class C = 2 _____

6. **HISTORY OF DISCIPLINARY** (clear conduct)
   Clear conduct for 36 months or more _____ -3
   Clear conduct for 24 months or more _____ -2
   Clear conduct for 12 months or more _____ -1

7. **PROGRAM/WORK PERFORMANCE** (last 6 months)
   Less than maximum good time ___ 0 Maximum good time _____ -2

8. **CURRENT AGE**
   21 and under _ 2 22 to 25 _ 1 26 to 34 _ 0 35 to 44 _ -1 45 and above _____ -2

9. **GANG MEMBERSHIP/ACTIVITIES IN THE PAST 10 YEARS**
   Yes _____ 2 No _____ 0

   TOTAL SCORE (Add 1 through 9)

Appendix B – Reasons for Overrides and Custody Cutoff Levels

Form CD-081200.3
Reviewed/Revised 03/31/22

NEW MEXICO CORRECTIONS DEPARTMENT
Custody Level Scale for Initial/Reclassification Scoring Forms

Male Inmates
- ☐ LEVEL IV — 10 points or higher on items 1-3, or 18 points and higher on items 1 through 8 (9)
- ☐ LEVEL III — 12-17 points on items 1 through 8 (9)
- ☐ LEVEL II — 6-11 points on items 1 through 8 (9)
- ☐ LEVEL I — 5 or less points on items 1 through 8 (9)

Female Inmates
- ☐ LEVEL IV — 10 points or higher on items 1-3, or 18 points and higher on items 1 through 8 (9)
- ☐ LEVEL III — 13-18 points on items 1 through 8 (9)
- ☐ LEVEL II — 7-12 points on items 1 through 8 (9)
- ☐ LEVEL I — 6 or less points on items 1 through 8 (9)

CUSTODY OVERRIDE FACTORS

MANDATORY OVERRIDE FACTORS: *Override factors are to be documented whether override is necessary or not, by checking the appropriate section*

Level I Facility Restrictions
- ____Has been convicted of a crime involving assaultive sexual conduct
- ____More than three years to projected release date
- ____ICE Detainer
- ____Has been convicted of a crime involving violence to a child
- ____Inconclusive NCIC
- ____Medical/Mental Health Restriction
- ____1st or 2nd Degree Murder

Level I and II Facility Restrictions
- ____More than five years to projected release date
- ____Current Offense for murder 1st degree
- ____Medical/Mental Health Restriction
- ____Felony Detainer/Active Felony Warrant
- ____Interstate Compact Inmate w/o custody reduction approval from sending state
- ____Escape from Level II facility within past 10 years
- ____Less than 12 months since PBMP release

Level I, II, and III Facility Restrictions (males only)
- ____Escape from a Level III or above
- ____Validated or Suspected STG Affiliation

DISCRETIONARY OVERRIDE TO INCREASE CUSTODY LEVEL: *Discretionary override factors are to be marked only if activated. (select most applicable)*
- ____Known behavior/management problems
- ____Suspected escape threat
- ____Escape from Level I
- ____Known gang affiliation/suspect
- ____Former Law Enforcement Officers
- ____Short time remaining to serve

DISCRETIONARY OVERRIDE TO LOWER CUSTODY LEVEL: *(select most applicable)*
- ____Inmate does not require level IV supervision
- ____Good institutional conduct
- ____Within six months projected remaining to serve

RECOMMENDED CUSTODY LEVEL
- ____Level I
- ____Level II
- ____Level III
- ____Level IV

RECOMMENDED HOUSING STATUS (male)
- ____General Population
- ____Special Management
- ____LCCF Unit 4
- ____MHTC
- ____LTCU

Gender Responsive
- ____Minimum
- ____Medium
- ____Maximum

Inmate's Signature:_____Date:_____

Classification Officer's Signature:_____Date:_____

Supervisor's Signature:_____Date:_____

Security Representative Signature:_____Date:_____

FINAL CUSTODY LEVEL
- ____Level I
- ____Level II
- ____Level III
- ____Level IV

FINAL HOUSING STATUS (male)
- ____General Population
- ____Special Management
- ____LCCF Unit 4
- ____MHTC
- ____LTCU

Gender Responsive
- ____Minimum
- ____Medium
- ____Maximum

Reason for Override:_____

Appendix C – Random Forest Model Results

Figure 1.
Variable Importance Plot for Random Forest for Initial Classification Tool (Male Inmates)

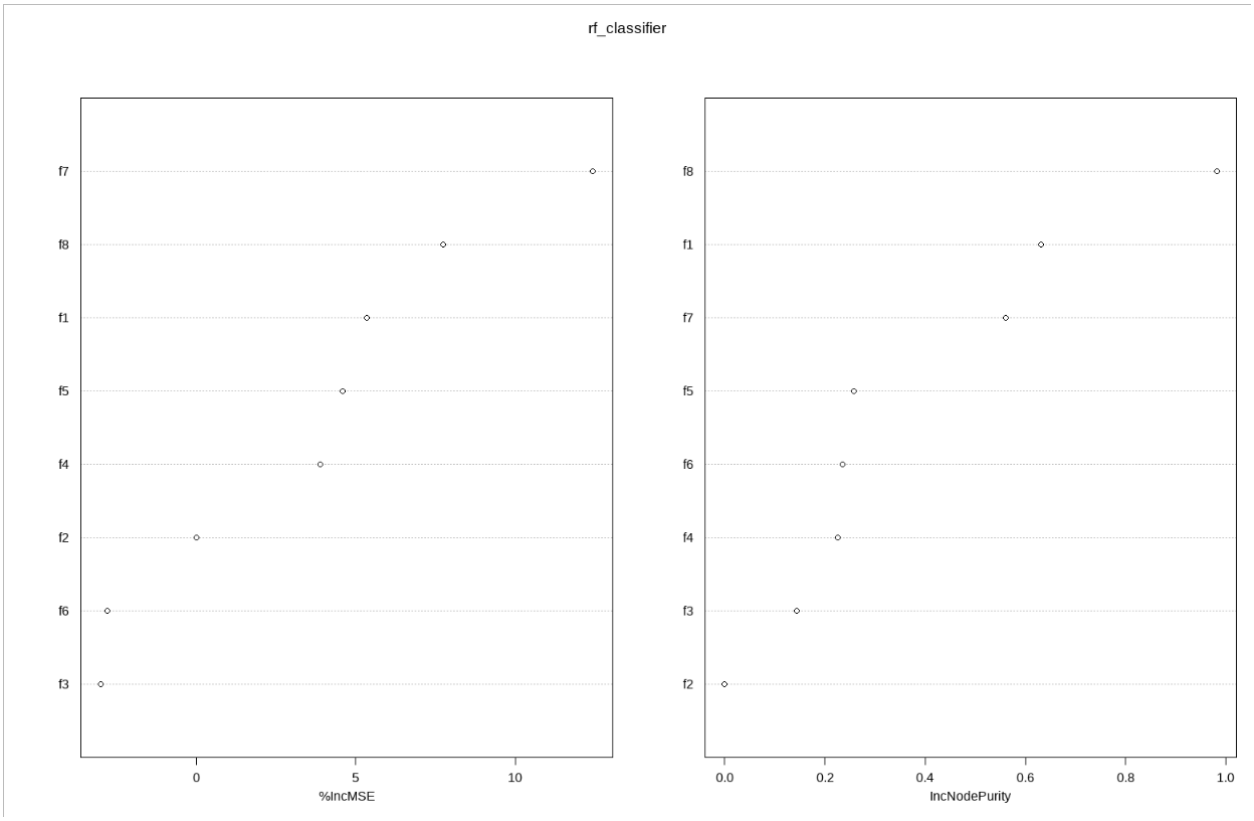


Figure 2.
Variable Importance Plot for Random Forest for Initial Classification Tool (Female Inmates)

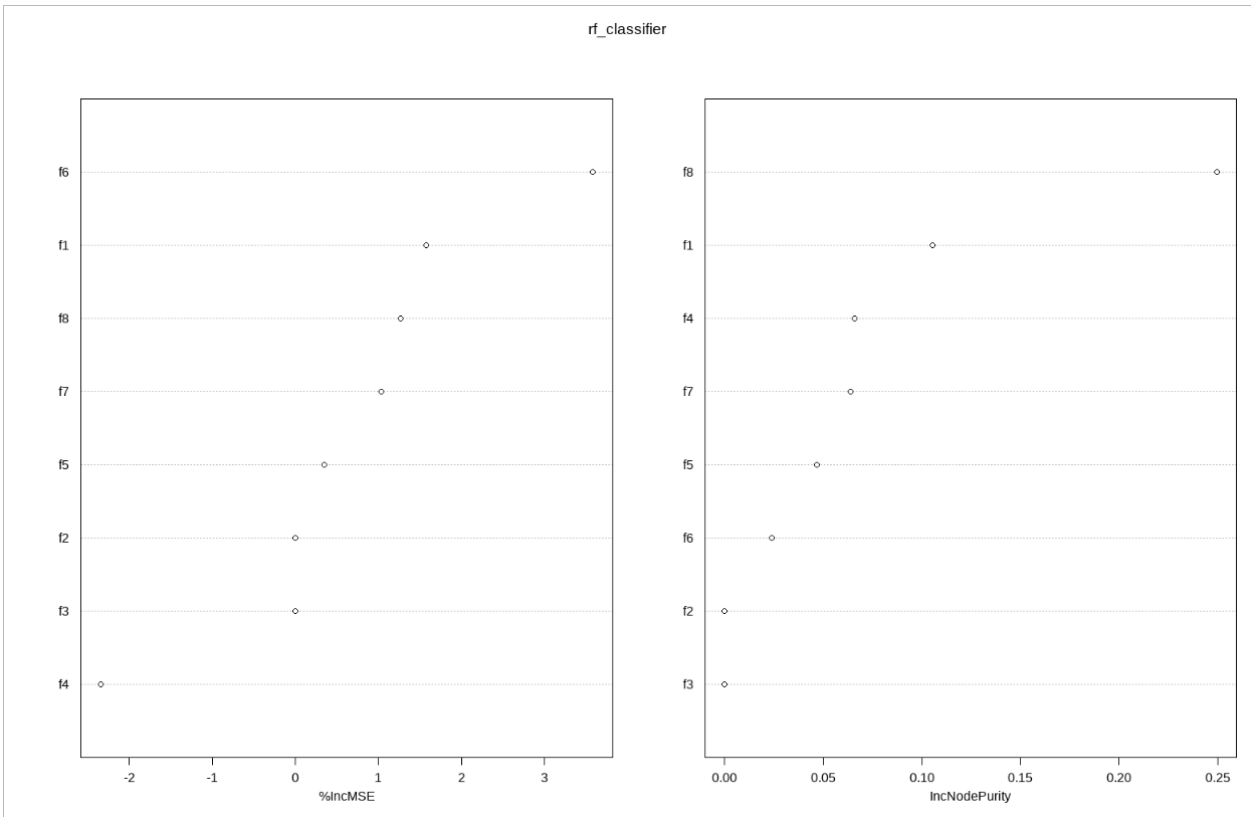


Figure 3.
Variable Importance Plot for Random Forest for Reclassification Tool (Male Inmates)

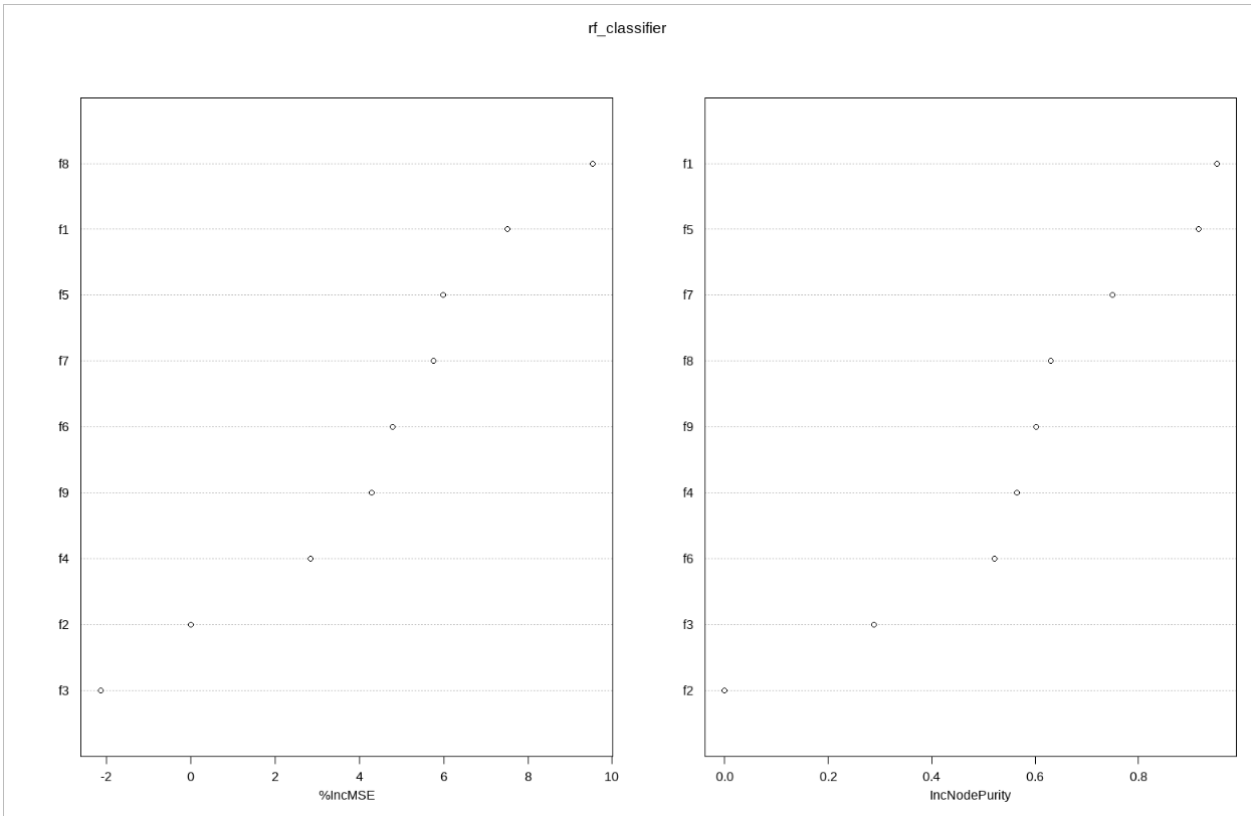
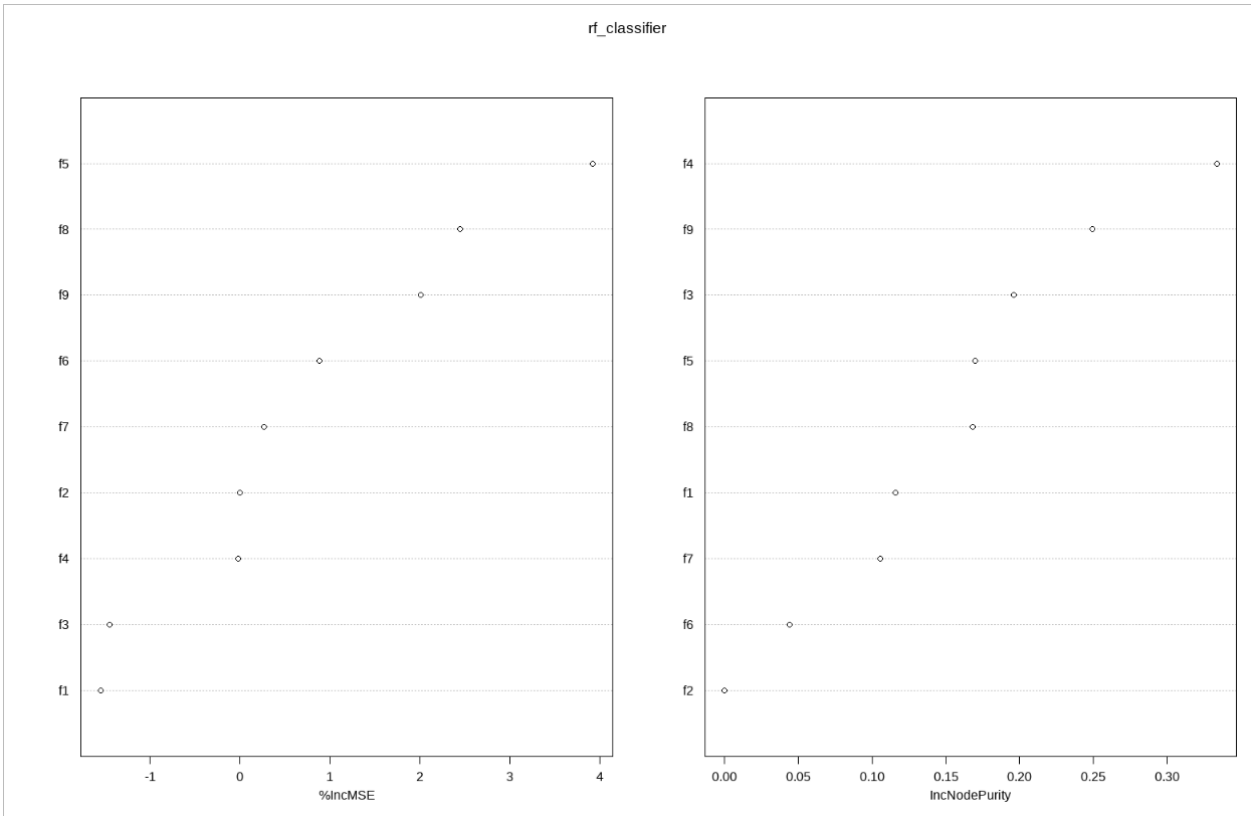


Figure 4.
Variable Importance Plot for Random Forest for Reclassification Tool (Female Inmates)



Appendix D – Empirical Justification For Retaining One Form for Male and Female Inmates and Changing Factor Scoring

In Figure 5, we present the percent of male and female inmates, within each custody level, who engaged in serious, violent misconduct (SVM) with bootstrapped 95% confidence intervals following initial classification and reclassification. Results generally suggest there were not significant differences in the percent of males and females engaging in SVM within each level, as revealed by the overlap in the error bars.

Figure 5.

Percent of Inmates Engaging in SVM Following Initial Classification by Level and Sex (n = 5,334)

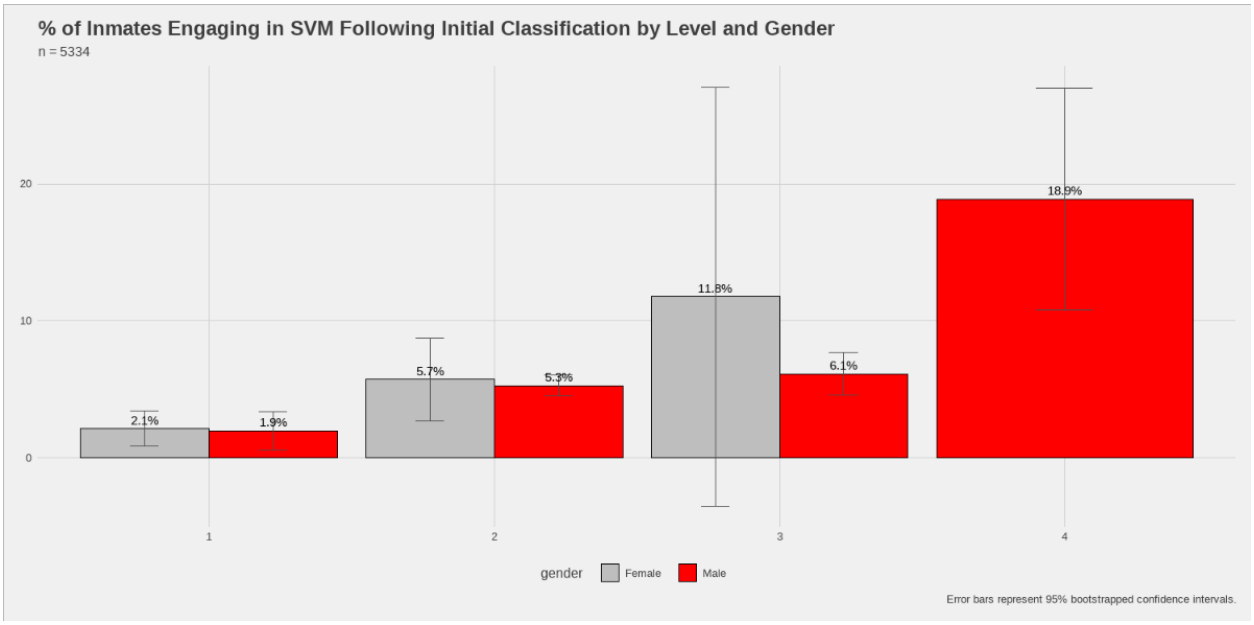
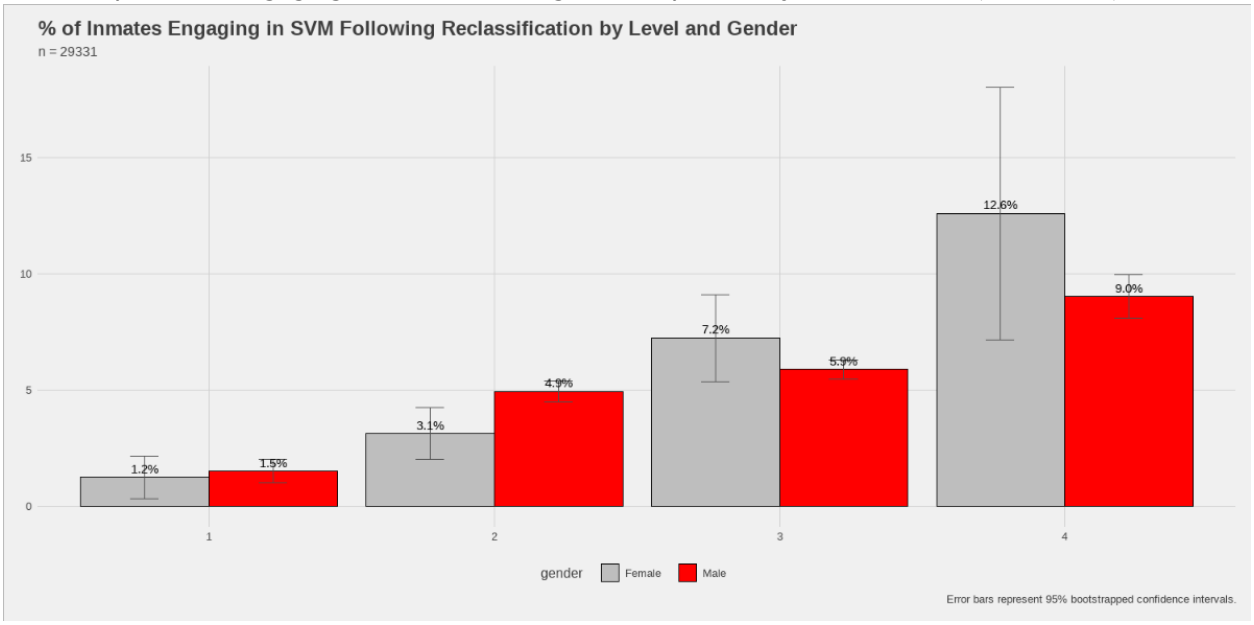


Figure 6.

Percent of Inmates Engaging in SVM Following Reclassification by Level and Sex (n = 29,331)



Figures 5 and 6 reveal (1) that the prevalence of SVM increases with custody level and (2) there are not statistically significant differences between male and female inmates in misconduct prevalence within levels.

In Figures 7-11, we present visualizations which compare the predicted probabilities that an inmate engaged in SVM by sex and factor score, controlling for other factor scores and custody levels, with 95%

confidence intervals[7]. Results of these models suggest that there were not statistically-significant differences in misconduct between male and female inmates within each factor item as given by the overlap of the 95% confidence intervals, and, while sex differences were more pronounced for the reclassification tool (i.e., on balance, female inmates had a substantively lower probability of SVM for some factors than male inmates), the difference did not rise to the threshold of statistical significance. Moreover, after reviewing predicted probabilities of SVM for male and female inmates, after controlling for custody level and other factors, males were not significantly more likely than females to engage in SVM, which may be attributable to the overall rarity of SVM. For these reasons, we believe there is empirical justification for retaining a single scoring form for male and female inmates instead of developing gender-specific scoring forms.

[7] Note that the qx_score language used within the visualization title corresponds to the relevant factor item on the current versions of the NMCD classification tools (e.g., q1_score refers to the "History of Institutional Misconduct" factor).

Figure 7.

Predicted Probabilities of SVM by Level and Sex for the Initial Classification Tool (Factors 1 – 4)

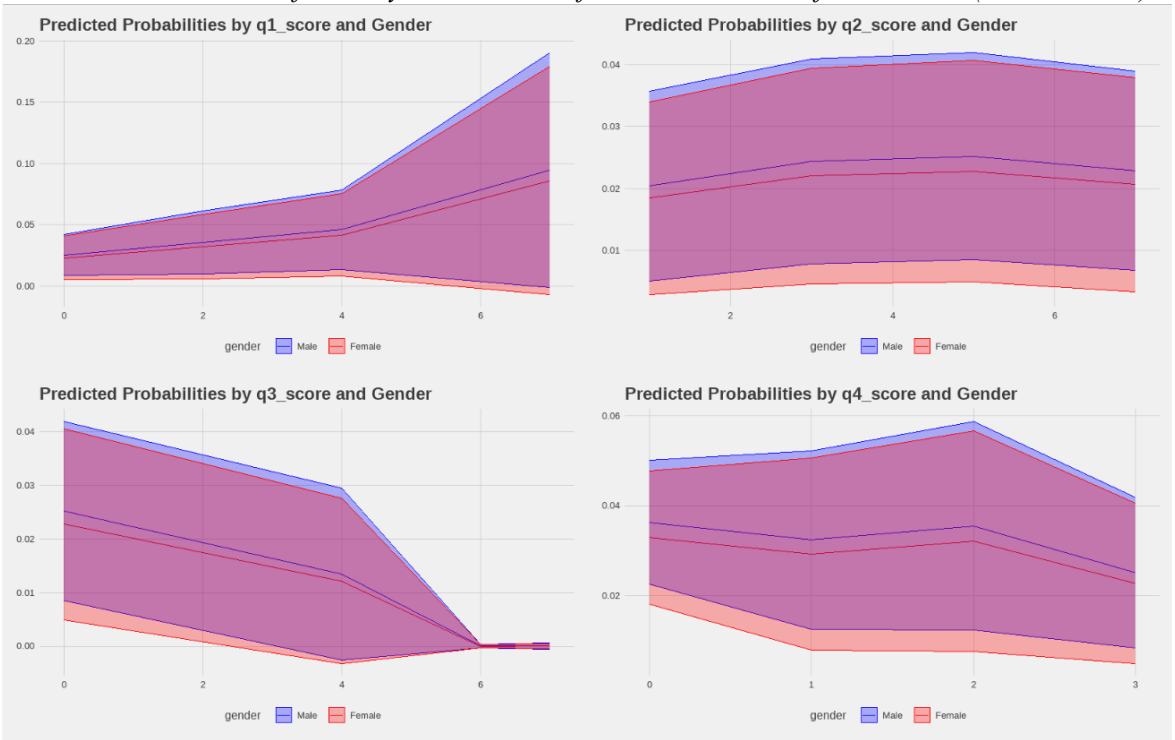


Figure 8.

Predicted Probabilities of SVM by Level and Sex for the Initial Classification Tool (Factors 5 – 8)

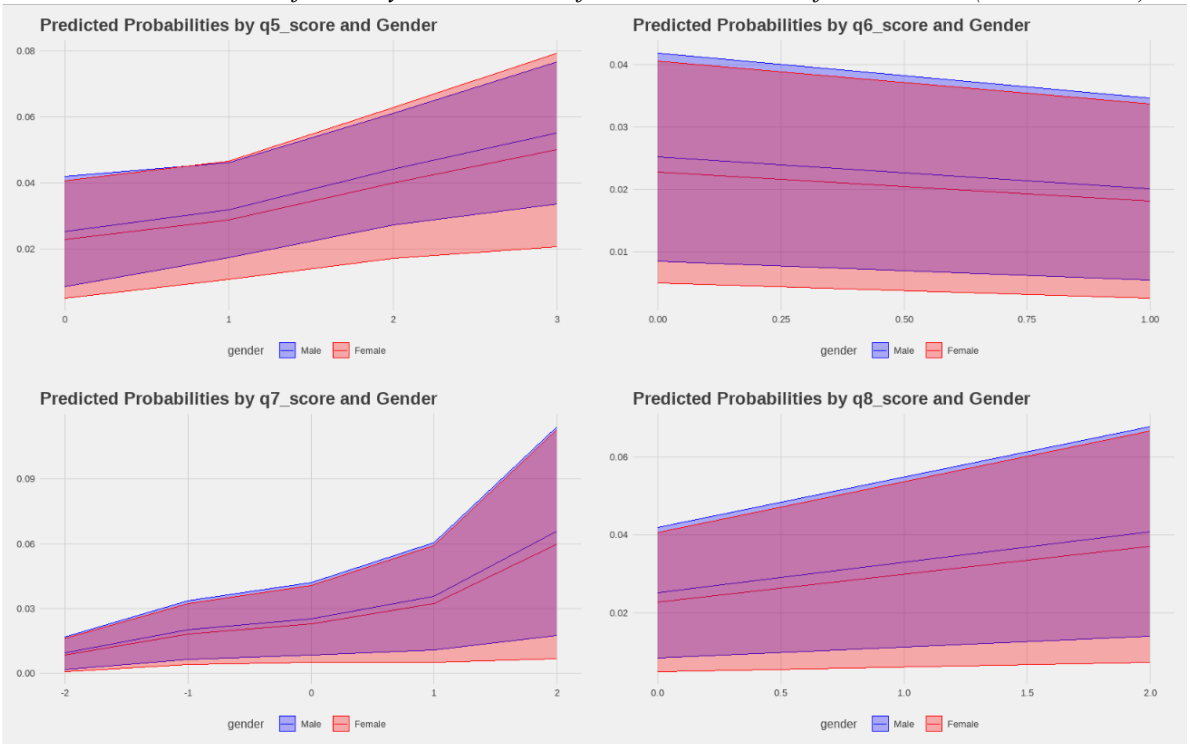


Figure 9.
Predicted Probabilities of SVM by Level and Sex for the Reclassification Tool (Factors 1 – 4)

Figure 10.
Predicted Probabilities of SVM by Level and Sex for the Reclassification Tool (Factors 5 – 8)

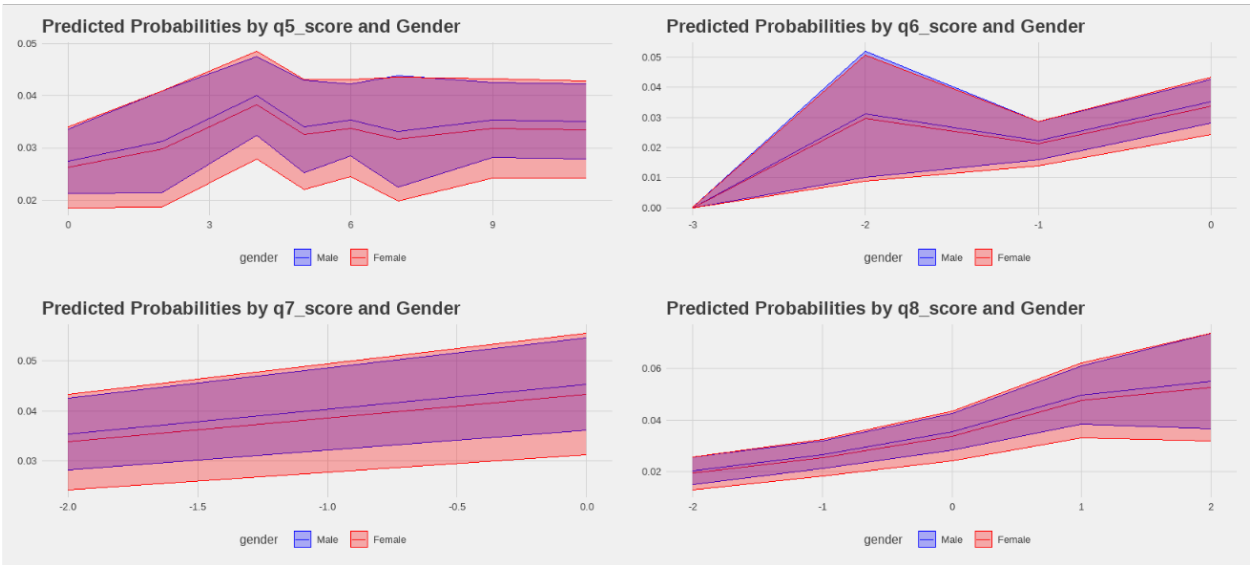
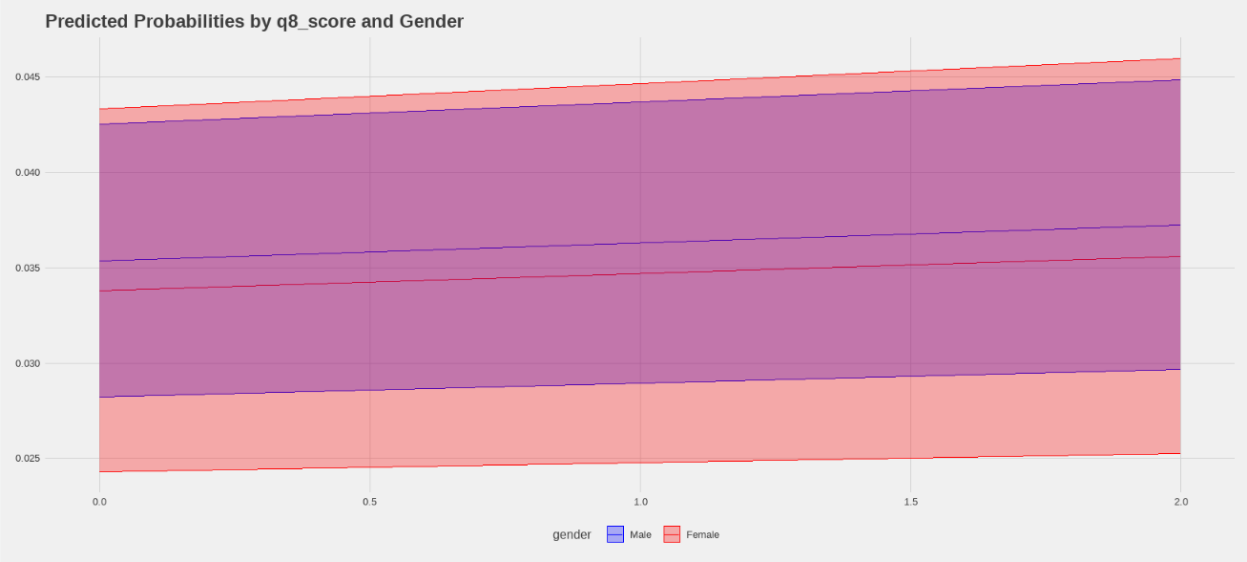


Figure 11.
Predicted Probabilities of SVM by Level and Sex for the Reclassification Tool (Factor 9)



Figures 7-11, in tandem with the predicted probabilities we present in Tables 1 – 4 which show the predicted probabilities and associated confidence intervals predicting both GM and SVM following classification, motivate proposed scoring changes to factors.

Table 1.

Predicted Probability of Engaging in GM by Factor with 95% Confidence Intervals [Initial Tool]

Factor	Predicted Probability (95% CI) - Male	Predicted Probability (95% CI) - Female
Institutional Violence (10 years)		
None (0)	36.9% (30.4-43.4)	37.7% (29.9-45.5)
10+ or Non-Violent (2)	51.4% (43.1-59.6)	52.2% (43.1-61.3)
Violent - No Weapon (4)	48.6% (40.2-57)	49.4% (40.1-58.8)
Violent – Weapon (7)	49.1% (30.7-67.6)	50% (31.1-68.9)
Current Conviction Risk		
Low (1)	38.2% (30.8-45.5)	39% (30.7-47.3)
Moderate (3)	40.1% (33.2-47)	40.9% (32.9-49)
High (5)	36.9% (30.4-43.4)	37.7% (29.9-45.5)
Highest (7)	38.5% (31.5-45.6)	39.4% (30.9-47.9)
Escape History (10 Years)		
None (0)	36.9% (30.4-43.4)	37.7% (29.9-45.5)
Escape from I/II (4)	32.4% (21.8-43)	33.1% (21.7-44.6)
Escape from III +(4)	100% (98.7-101.3)	100% (98.7-101.3)
Escape with Violence (7)	0% (-0.3-0.3)	0% (-0.3-0.3)
Prior Felony Convictions		
None (0)	34.9% (30.9-38.9)	35.7% (31.3-40)
One (1)	36.9% (30.9-42.9)	37.7% (30.5-44.9)
Two (2)	38% (31.2-44.7)	38.8% (30.9-46.7)
Three+ (3)	36.9% (30.4-43.4)	37.7% (29.9-45.5)
Risk Prior Convictions		
None/Low (0)	36.9% (30.4-43.4)	37.7% (29.9-45.5)
Moderate (1)	40.3% (35.4-45.2)	41.1% (34.5-47.8)
High (2)	40.3% (35.9-44.8)	41.2% (34.8-47.5)
Highest (3)	44.3% (39.4-49.2)	45.2% (38.3-52)
Alcohol/Drug Abuse		
None (0)	36.9% (30.4-43.4)	37.7% (29.9-45.5)
Trafficking (1)	33.9% (27-40.9)	34.7% (26.4-43.1)
Current Age		
<=21 (2)	66.7% (58.3-75.1)	67.5% (57.8-77.2)
22 – 25 (1)	51.3% (43.5-59)	52.1% (42.8-61.5)
26-34 (0)	36.9% (30.4-43.4)	37.7% (29.9-45.5)
35-44 (-1)	24.6% (19.3-29.8)	25.2% (19-31.4)
45+ (-2)	17.6% (13.2-22.1)	18.1% (12.9-23.3)
Gang Membership (10 Years)		
No (0)	36.9% (30.4-43.4)	37.7% (29.9-45.5)
Yes (2)	50% (42.8-57.3)	50.9% (42.1-59.6)

Table 2.

Predicted Probability of Engaging in SVM by Factor with 95% Confidence Intervals [Initial Tool]

Factor	Predicted Probability (95% CI) - Male	Predicted Probability (95% CI) - Female
Institutional Violence (10 years)		
None (0)	2.5% (0.9-4.2)	2.3% (0.5-4.1)
10+ or Non-Violent (2)	3.5% (1-6.1)	3.2% (0.6-5.8)
Violent - No Weapon (4)	4.6% (1.4-7.8)	4.2% (0.8-7.5)
Violent – Weapon (7)	9.5% (-0.1-19)	8.6% (-0.7-17.9)
Current Conviction Risk		
Low (1)	2% (0.5-3.6)	1.8% (0.3-3.4)
Moderate (3)	2.4% (0.8-4.1)	2.2% (0.5-3.9)
High (5)	2.5% (0.9-4.2)	2.3% (0.5-4.1)
Highest (7)	2.3% (0.7-3.9)	2.1% (0.3-3.8)
Escape History (10 Years)		
None (0)	2.5% (0.9-4.2)	2.3% (0.5-4.1)
Escape from I/II (4)	1.3% (-0.3-3)	1.2% (-0.3-2.8)
Escape from III +(4)	0% (0-0)	0% (0-0)
Escape with Violence (7)	0% (-0.1-0.1)	0% (0-0)
Prior Felony Convictions		
None (0)	3.6% (2.3-5)	3.3% (1.8-4.8)
One (1)	3.2% (1.3-5.2)	2.9% (0.8-5.1)
Two (2)	3.6% (1.2-5.9)	3.2% (0.8-5.7)
Three+ (3)	2.5% (0.9-4.2)	2.3% (0.5-4.1)
Risk Prior Convictions		
None/Low (0)	2.5% (0.9-4.2)	2.3% (0.5-4.1)
Moderate (1)	3.2% (1.7-4.6)	2.9% (1.1-4.7)
High (2)	4.4% (2.7-6.1)	4% (1.7-6.3)
Highest (3)	5.5% (3.4-7.7)	5% (2.1-7.9)
Alcohol/Drug Abuse		
None (0)	2.5% (0.9-4.2)	2.3% (0.5-4.1)
Trafficking (1)	2.0% (0.5-3.5)	1.8% (0.2-3.4)
Current Age		
<=21 (2)	6.6% (1.8-11.4)	6% (0.7-11.3)
22 – 25 (1)	3.6% (1.1-6)	3.2% (0.5-5.9)
26-34 (0)	2.5% (0.9-4.2)	2.3% (0.5-4.1)
35-44 (-1)	2.0% (0.7-3.4)	1.8% (0.4-3.2)
45+ (-2)	0.9% (0.2-1.7)	0.8% (0.1-1.6)
Gang Membership (10 Years)		
No (0)	2.5% (0.9-4.2)	2.3% (0.5-4.1)
Yes (2)	4.1% (1.4-6.8)	3.7% (0.7-6.7)

Table 3.

Predicted Probability of Engaging in GM by Factor with 95% Confidence Intervals [Reclass Tool]

Factor	Predicted Probability (95% CI) - Male	Predicted Probability (95% CI) - Female
Institutional Violence (10 years)		
None (0)	40.9% (38.6-43.2)	43.1% (39.9-46.3)
10+ or Non-Violent (1)	45.8% (43.6-48)	48% (44.9-51.2)
Violent - No Weapon (3)	44.5% (42.4-46.7)	46.7% (43.6-49.9)
Violent – Weapon (6)	39.3% (35.3-43.2)	41.4% (36.8-45.9)
Current Conviction Risk		
Low (1)	40.4% (37.5-43.3)	42.6% (39-46.2)
Moderate (2)	43.6% (41.1-46.1)	45.8% (42.5-49.1)
High (4)	40.9% (38.6-43.2)	43.1% (39.9-46.3)
Highest (6)	39.7% (37.2-42.2)	41.8% (38.4-45.3)
Escape History (10 Years)		
None (0)	40.9% (38.6-43.2)	43.1% (39.9-46.3)
Escape from I/II (3)	38.4% (34.3-42.4)	40.5% (35.8-45.1)
Escape from III +(5)	22.4% (-1.7-46.5)	23.9% (-1.4-49.3)
Escape with Violence (6)	37% (-19.6-93.6)	39.1% (-18.7-96.9)
Prior Felony Convictions		
None (0)	41.3% (39.1-43.5)	43.4% (40.6-46.3)
One (1)	39.2% (36.7-41.7)	41.3% (38-44.6)
Two (2)	41.3% (38.7-43.9)	43.5% (40.1-46.9)
Three+ (3)	40.9% (38.6-43.2)	43.1% (39.9-46.3)
Disciplinary Convictions Received		
0	34.2% (32-36.3)	36.2% (33.2-39.1)
2	43.2% (39.8-46.5)	45.4% (41.3-49.4)
4	42.1% (40-44.3)	44.3% (41.2-47.4)
5	40.1% (37.2-43.1)	42.3% (38.6-45.9)
6	49.7% (47.4-52)	51.9% (48.8-55.1)
7	41.9% (38.1-45.6)	44% (39.7-48.4)
9	40.9% (38.6-43.2)	43.1% (39.9-46.3)
11	43.1% (40.8-45.5)	45.3% (42.1-48.5)
History of Disciplinary		
Clear Conduct 36+ Months (-3)	20% (12.4-27.7)	21.5% (13.3-29.7)
Clear Conduct 24+ Months (-2)	28.7% (22.6-34.7)	30.5% (24-37.1)
Clear Conduct 12+ Months (-1)	30.5% (27.9-33.1)	32.4% (29.1-35.7)
No Clear Conduct (0)	40.9% (38.6-43.2)	43.1% (39.9-46.3)
Program/Work Experience		
Maximum Good Time (-2)	40.9% (38.6-43.2)	43.1% (39.9-46.3)
Less Than Maximum Good Time (0)	46% (43.6-48.5)	48.2% (44.9-51.5)
Current Age		
<=21 (2)	51% (46.3-55.7)	53.2% (48-58.5)
22 – 25 (1)	47.8% (45-50.7)	50%(46.3-53.8)
26-34 (0)	40.9% (38.6-43.2)	43.1%(39.9-46.3)
35-44 (-1)	33.3% (31.3-35.4)	35.3%(32.4-38.2)
45+ (-2)	22.9% (20.9-24.9)	24.5%(21.8-27.2)

Gang Membership (10 Years)		
No (0)	40.9% (38.6-43.2)	43.1%(39.9-46.3)
Yes (2)	45.2% (42.8-47.6)	47.4% (44-50.8)

Table 4.

Predicted Probability of Engaging in SVM by Factor with 95% Confidence Intervals [Reclass Tool]

Factor	**Predicted Probability (95% CI) - Male**	**Predicted Probability (95% CI) - Female**
Institutional Violence (10 years)		
None (0)	3.5% (2.8-4.3)	3.4% (2.4-4.3)
10+ or Non-Violent (1)	4.6% (3.8-5.4)	4.4% (3.2-5.5)
Violent - No Weapon (3)	7.8% (6.5-9)	7.4% (5.6-9.3)
Violent – Weapon (6)	8.8% (6.5-11.1)	8.4% (5.7-11.2)
Current Conviction Risk		
Low (1)	3.4% (2.5-4.3)	3.3% (2.2-4.3)
Moderate (2)	3.7% (2.9-4.5)	3.5% (2.5-4.6)
High (4)	3.5% (2.8-4.3)	3.4% (2.4-4.3)
Highest (6)	4.2% (3.3-5.1)	4% (2.8-5.2)
Escape History (10 Years)		
None (0)	3.5% (2.8-4.3)	3.4% (2.4-4.3)
Escape from I/II (3)	4.2% (2.8-5.6)	4% (2.5-5.6)
Escape from III +(5)	5.3% (-5.2-15.8)	5.1% (-5-15.1)
Escape with Violence (6)	37.5% (-20.1-95.1)	36.4% (-20.8-93.5)
Prior Felony Convictions		
None (0)	3.2% (2.6-3.9)	3.1% (2.3-3.9)
One (1)	3.3% (2.5-4)	3.1% (2.2-4.1)
Two (2)	4.2% (3.3-5.1)	4% (2.9-5.2)
Three+ (3)	3.5% (2.8-4.3)	3.4% (2.4-4.3)
Disciplinary Convictions Received		
0	2.7% (2.1-3.4)	2.6% (1.9-3.4)
2	3.1% (2.1-4.1)	3% (1.9-4.1)
4	4% (3.2-4.7)	3.8% (2.8-4.9)
5	3.4% (2.5-4.3)	3.3% (2.2-4.3)
6	3.5% (2.8-4.2)	3.4% (2.4-4.3)
7	3.3% (2.3-4.4)	3.2% (2-4.4)
9	3.5% (2.8-4.3)	3.4% (2.4-4.3)
11	3.5% (2.8-4.2)	3.3(2.4-4.3)
History of Disciplinary		
Clear Conduct 36+ Months (-3)	0% (0-0)	0% (0 – 0)
Clear Conduct 24+ Months (-2)	3.1% (1 - 5.2)	3.0% (0 – 5.1)
Clear Conduct 12+ Months (-1)	2.2% (1.6- 2.9)	2.1% (1 – 2.9)
No Clear Conduct (0)	3.5% (2.8 - 4.3)	3.4% (2.4 – 4.3)
Program/Work Experience		
Maximum Good Time (-2)	3.5% (2.8-4.3)	3.4% (2.4-4.3)
Less Than Maximum Good Time (0)	4.5% (3.6-5.5)	4.3% (3.1-5.6)
Current Age		
<=21 (2)	5.5% (3.7-7.4)	5.3% (3.2-7.4)
22 – 25 (1)	5% (3.8-6.1)	4.7% (3.3-6.2)
26-34 (0)	3.5% (2.8-4.3)	3.4% (2.4-4.3)
35-44 (-1)	2.7% (2.1-3.2)	2.5% (1.8-3.2)
45+ (-2)	2% (1.5-2.6)	1.9% (1.3-2.6)
Gang Membership (10 Years)		

No (0)	3.5% (2.8-4.3)	3.4% (2.4-4.3)
Yes (2)	3.7% (3-4.5)	3.6% (2.5-4.6)

Consider the scoring for age relative to the number and type of disciplinary convictions received on the reclassification tool. On the current version of the reclassification tool, a male inmate who committed a Class A, Class B, and Class infraction in the prior 24 months had a 3.5% predicted probability of engaging in SVM in the following six months and a 43.1% predicted probability of engaging in GM; a male inmate with the minimum score of 0 on this factor, indicative of committing no Class A, B, or C offenses in the prior 24 months would have a 2.7% predicted probability of engaging in SVM and a male inmate with the maximum score of 11 on this factor, indicative of committing Class A, B, and C offenses in the prior 24 months, would have a 3.5% predicted probability of engaging in SVM, an increase of less than 1 percentage point risk. If an inmate committed a Class A, Class B, and Class C infraction in the 24 months before they were reclassified, they would receive 11 points on this one factor alone.

However, on the current version of the reclassification tool, a male inmate who was 21 years old or younger received 2 points on the tool yet had a 5.5% predicted probability of engaging in SVM and a 51% probability of engaging in GM, predicted probabilities which are both greater than the predicted probability of SVM and GM associated with scoring an 11 on the disciplinary convictions received factor. Additionally, relative to a male inmate aged 45+, a male inmate age 21 or under had more than double the predicted probability of engaging in both GM and SVM (51% versus 21.9%; 5.5% versus 2%). Moreover, per the variable importance plots in Appendix C, age emerged as either the first or second most important factors for predicting SVM for both male and female inmates.

Because of this, we reasoned that it would make sense to weight a factor, like younger age, more heavily relative to other factors on the tool, which is why the maximum value for the factor increased from 2 to 8 and the relative values assigned to committing Class A, B, and C infractions on the reclassification tool were reduced by a point each.

Similarly, on the current reclassification tool, inmates who had the highest “Risk of Current Conviction” score were assigned the same number of points, 6, as inmates with a history of institutional violence with a weapon, even though inmates with a history of institutional violence had considerably higher odds of engaging in SVM than inmates who had the highest risk of current conviction. Given the similarities in predicted probabilities between inmates who had a violent incident with and without a weapon, we recommend scoring these factors similarly, as the use of weapon was immaterial for predicting either heightened GM or SVM risk.

Appendix E – Empirical Justification for Reducing Lookback Timeframes for History of Institutional Violence Factor

A 2020 report by the New Mexico Legislative Finance Committee Program Evaluation Unit noted the following about the NMCD's current classification tools:

> *Disciplinary incidents currently follow inmates for a decade, with points added for serious misconduct within the last 10 years. But the literature review ISR conducted in 2017 found research suggesting only recent misconduct – over the past year – is predictive of risk for future misconduct. While the reclassification form allows points to be deducted for recent good behavior, with higher deductions for longer misconduct-free stretches, those deductions could be canceled out by points for old offenses. Revising this aspect of the scoring instrument may make it easier for inmates to access less restrictive custody levels without compromising safety.*
>
> *ISR's proposed revision to this scoring factor in the tool it tested in the 2019 report considers disciplinary history only over the past year. It should be noted, however, that the scoring system in use prior to the Santa Rosa riots also only considered institutional misconduct for a year. The independent investigation concluded this timeline was too short for New Mexico's inmate population and contributed to underclassification of violent inmates and gang members. The forthcoming validation study to be completed by ISR by summer 2022 should attempt to determine whether one year is appropriate for the current population, or whether something more conservative is warranted.* (New Mexico Legislative Finance Committee 2020, pg. 15)

More recent past behavior is, generally, a better predictor of future behavior than more distant past behavior (Loughran et al., 2017; Albarracin & Wyer 2000; Ouellette & Wood 1998). On this assumption and empirical grounding, we recommend that the NMCD reduce the lookback timeframes embedded within tool factors which involve lookback periods (e.g., history of institutional violence; escape history; gang membership).

There are some practical concerns associated with using a 10-year lookback window for the history of institutional misconduct factor. The further back in time an inmate's misconduct history extends, the less relevant the prior misconduct becomes for predicting future behavior. Inmates may have evolved over a long period and an inmate's behaviors and attitudes may no longer reflect past actions.

To the extent that prisons serve rehabilitation purposes, the use of a 10-year lookback period may undermine rehabilitative goals. Inmates who make significant strides in their personal growth and behavior improvement may find it discouraging if their past misconduct from a distant period continues to heavily influence their classification. Problematically, the point penalty from behaviors far in the past could disincentivize good behavior since the past behavior is disproportionately weighted in relation to its predictive power. Balancing the need for accurate risk assessment with considerations of fairness, rehabilitation, and practicality is crucial when determining the appropriate time frame for evaluating an inmate's history of institutional misconduct.

To empirically evaluate the relationship between time of prior misconduct and future misconduct (i.e., the first factor in both the initial and reclassification forms), we used a logistic model to predict whether any misconduct (GM) occurred in the six months following reclassification for male and female inmates after

statistically controlling for age, Sex, race-ethnicity, custody tool level, custody tool score, counts of any guilty misconduct events binned by year [i.e., five dummy variables which, within each inmate, counted the number of guilty misconduct infractions in that year window (e.g., counts of misconduct in Year 1 prior to reclass; counts of misconduct in Year 2 prior to reclass through Year 5)], and a variable which, within each inmate, averaged the typical time between all guilty misconduct events and reclassification.

We then plotted predicted probabilities inmates engaged in misconduct based on the average time between misconduct and the reclassification date variable, which generated Figure 1 below, which shows that, after controlling for other factors, the predicted probability that an inmate engaged in misconduct significantly increased as the average within-inmate time between misconduct and reclassification increased (e.g., inmates who had an average of eight years between their misconduct and reclassification had about a 50% probability of engaging in misconduct following reclass in contrast to individuals who had an average of two years between their prior misconducts and reclassification, which had about a 85% predicted probability of engaging in misconduct following reclass).

Additionally, misconduct counts which occurred in each year prior to the reclassification were independently predictive of misconduct with counts of misconduct in the year prior to reclassification being the mostly strongly predictive of post-reclass misconduct relative to counts in other years. We also ran a series of random forest models including dummy variables which indicate whether an inmate engaged in any general misconduct (GM) within the year prior to the reclass event (i.e., *onep_indicator* in the variable importance plot in Figure 2), two years prior to the reclassification(i.e., *twop_indicator*), through five years prior to reclassification, alongside controls for custody level, tool score, and age. The result of the random forest specification seems to similarly suggest that more recent misconduct is more important for predicting post-reclass misconduct than more temporally-distant misconduct (e.g., *one_p indicator* > *threep_indicator/two_p indicator* > *fivep_indicator* and *fourp_indicator*).

While the choice of where to draw the lookback threshold is not guided by hard-and-fast rules, we advise that using a three-year lookback period would be an appropriate middle-ground between the NMCD's current practice of using 10 years of misconduct history, comments from NMCD staff about the practicability of shortening the timeframe, and the results of the logistic and random forest models.

Figure 12.

Predicted Probability of Engaging in GM Based on Average Time of Misconduct

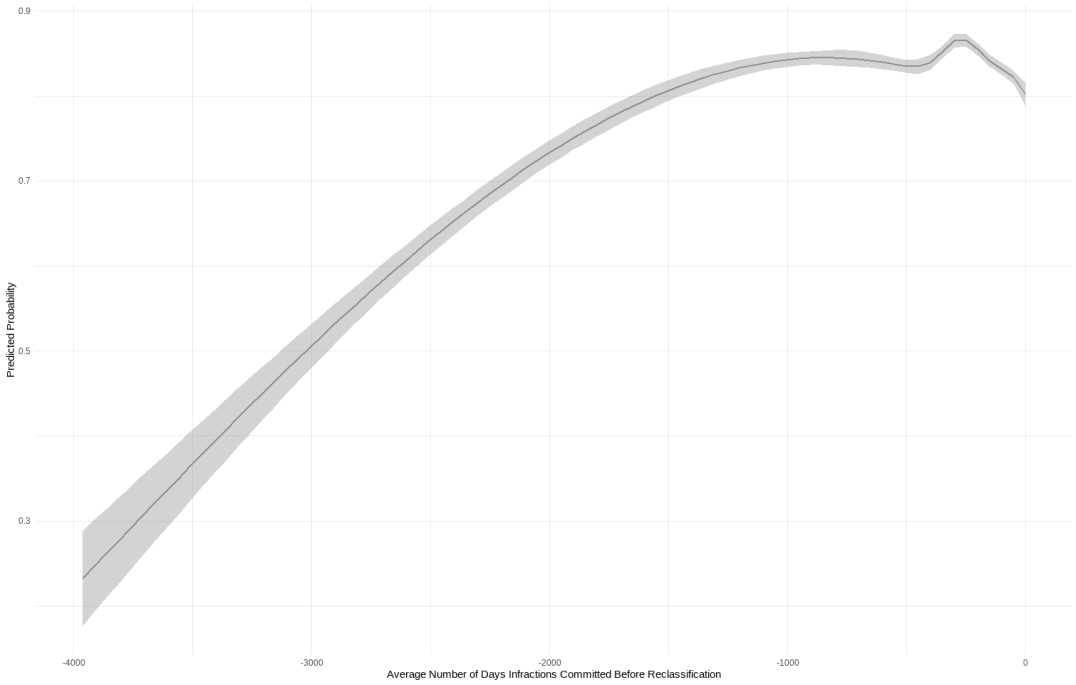
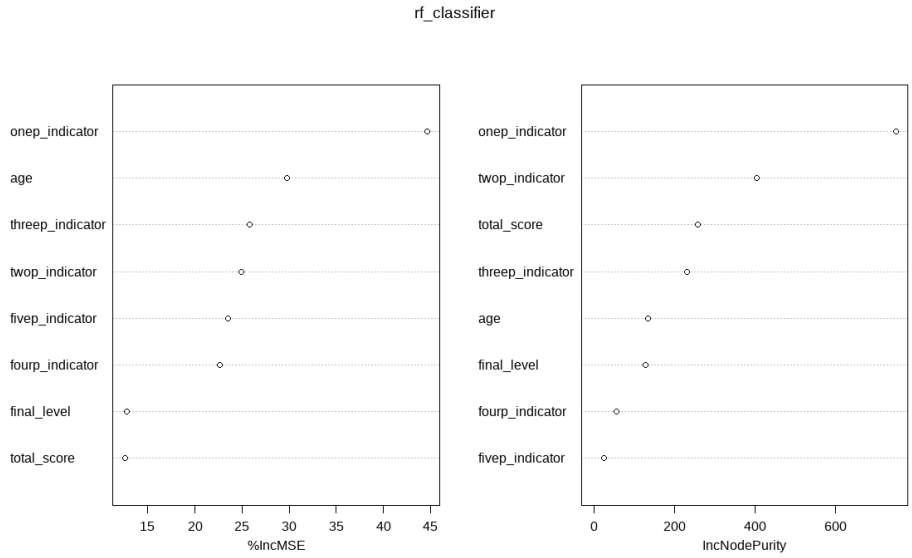


Figure 13.

Random Forest VIPs of Average Time to Misconduct on Post-Reclass Misconduct Counts



Appendix F – Proposed Revisions to Existing NMCD Classification Tools

Proposed Scoring Form Changes: Initial Classification Tool

Form CD-081200.1
ISR Revisions

NEW MEXICO CORRECTIONS DEPARTMENT
INITIAL CUSTODY SCORING FORM

Inmate's Name:_____NMCD#_____
Last *First* *MI*

Classification Officer:_____Classification Date:_____

1. **HISTORY OF INSTITUTIONAL ADJUSTMENT/VIOLENCE.** (Review individual's entire background for 3 years prior to classification date to include juvenile incidents) (Include date of incident; rate most severe)
   None _____ 0
   Ten or more non-violent disciplinary reports _____ 2
   Non-Violent /Serious Class A Level Incidents _____ 2
   Violent Incident with no weapon, serious injury or death _____ 6
   Violent Incident involving a weapon, serious injury or death _____ 8

2. **CURRENT CONVICTION SEVERITY** (score the most serious conviction, list offense and date)
   Low _____ 0
   Moderate _____ 1
   High _____ 2
   Highest 3

3. **ESCAPE HISTORY** (Last 3 years from this rating date. List date of escape)
   None _____ 0
   Escape/Attempted Escape from Level I or II, County Jail, Juvenile Facility, or Peace Officer (no violence) _____ 3
   Escape/Attempted Escape from Level III facility or above (no violence) 5
   Escape/Attempted Escape (with violence) _____ 10

4. **PRIOR # OF FELONY CONVICTIONS** (Do not include current conviction; list offenses and dates.) None _ 0 One or More _____ 1

5. **PRIOR CONVICTION SEVERITY** (Score the most serious offence; list offence and date) None/Low _____ 0 Moderate _____ 2 High ___ 4 Highest _ 6

6. **CURRENT AGE**
   21 and under _ 8 22 to 25 _ 5 26 to 34 _ 4 35 to 44 _ 2 45 and above _____ 0

7. **GANG MEMBERSHIP/ACTIVITIES IN THE PAST 3 YEARS**
   Yes _____ 3 No _____ 0

TOTAL SCORE (Add 1 through 7)

Level Thresholds:

- Level I: 0 - 8
- Level II: 9 - 14
- Level III: 15 - 20
- Level IV: 21+

Proposed Scoring Form Changes: Reclassification Tool

Form CD-081200.2
ISR Revisions

NEW MEXICO CORRECTIONS DEPARTMENT
RECLASSIFICATION SCORING FORM

Inmate's Name:_____NMCD#_____
Last *First* *MI*

Classification Officer:_____Reclassification Date:_____

1. **HISTORY OF INSTITUTIONAL ADJUSTMENT/VIOLENCE.** (Review individual's entire background for 3 years prior to classification date to include juvenile incidents) (Include date of incident; rate most severe)
   None _____ 0
   Non-Violent /Serious Class A Level Incidents _____ 3
   Violent Incident with no weapon, serious injury or death_____ 8
   Violent Incident involving a weapon, serious injury or death_____ 10

2. **CURRENT CONVICTION SEVERITY** (score the most serious conviction, list offense and date)
   Low_____ 0
   Moderate _____ 1
   High_____ 2
   Highest _____ 3

3. **ESCAPE HISTORY** (Last 3 years from this rating date. List date of escape)
   None _____ 0
   Escape/Attempted Escape from Level I or II, County Jail, Juvenile Facility, or Peace Officer (no violence) _____ 3
   Escape/Attempted Escape from Level III facility or above (no violence) _____ 6
   Escape/Attempted Escape (with violence) _____ 10

4. **PRIOR # OF FELONY CONVICTIONS** (Do not include current conviction; list offenses and dates.) None _ 0 One or more _____ 1

5. **DISCIPLINARY CONVICTIONS RECEIVED** (only one per class)
   Last 24 months: One or more class A = 4 _____
   Last 24 months: One or more class B = 3 _____
   Last 24 months: One or more class C = 2_____

6. **CURRENT AGE**
   21 and under _ 8 22 to 25 _ 7 26 to 34 _ 3 35 to 44 _ 2 45 and above _____ 0

7. **GANG MEMBERSHIP/ACTIVITIES IN THE PAST 3 YEARS**
   Yes _____ 2 No _____ 0

TOTAL SCORE (Add 1 through 7)

LEVEL THESHOLDS:

- Level I: 0 - 8
- Level II: 9 - 14
- Level III: 15 - 20
- Level IV: 21+