8-2016

Development and Validation of the Nebraska Department of Correctional Services Prison Classification System

Zachary Hamilton

Alex Kigerl

Development and Validation of the Nebraska Department of Correctional Services Prison Classification System

Zachary Hamilton, Ph.D.

&

Alex Kigerl, Ph.D.

August, 2016

This research was funded by:

The Nebraska Center for Justice Research

University of Nebraska Omaha

6001 Dodge Street

Table of Contents

EXECUTIVE SUMMARY

Over the last 45 years, the Nebraska Department of Correctional Services (NDCS) has made two substantial changes to its classification system. The first system was created and implemented in the 1970s. In 2005, Patricia Hardyman was contracted to update and modify the classification and reclassification system. Using statistical analyses of available data, a set of prediction models were created to score inmates on items that predicted future infraction behavior. However, the primary issues of the tools were that inmates' scores were routinely over-classified and that substantial uses of overrides (approximately 40%) were indicated. After a review of the tools' development methods, several issues were identified, including: a small development sample, a lack of prediction and outcome item specificity, lack of item weighting, a less than optimal feature select strategy, inefficient validation techniques, and the tool's inability to assess infraction prediction post-transfer. It was therefore determined that this system required major improvements and a research project was outlined and contracted by the University of Nebraska, Omaha.

The project was outlined to complete five major project stages. First, a process evaluation was completed consisting of an NDCS document review, classification and reclassification was observed, key staff were interviewed, focus groups of staff, research, and administrative personnel were facilitated to assess key advantages, disadvantages, and potential areas for improvement of the current classification system and processes. Several themes were described around the tool's functionality and use, including: extensive use of overrides, bed-space driven placement needs, a lack of specialized placement options, mandatory override restrictions, illogical inclusion of certain predictor items, inconsistent scoring across assessment types, and programming availability. The culmination of findings outlined potential issues impacting accuracy and usage of scoring from users as well as methodological limitations of current classification and reclassification model construction.

Next, we reviewed available NDCS data and developed an analysis plan for tool construction. Working with NDCS research staff, we identified a sample frame for initial and reclassification analyses. Ultimately, the samples collected consisted of 9,072 male and 1,582 female initial classification and 35,098 male and 2,449 female reclassification assessments of offenders incarcerated and supervised by the NDCS during the study period of August, 1991 and June, 2015.

We then sought to develop statistical models for infraction prediction models. First, the current classification and reclassification models were assessed for their ability to predict infraction behavior. Next, feature selection procedures were completed, selecting items that improved prediction of three infractions outcomes – violent, serious, and non-serious[1]. This was completed using several advanced multivariate selection techniques. These models were created in an effort to improve upon, and replace, the current classification and reclassification models.

Study Findings

Validation procedures were then completed. The new models created were compared to those currently in use based on the industry standard statistic – the Area Under the Curve (AUC). New models demonstrated substantial improvement compared to the previously developed Haryman tools. These findings confirm the predictive improvements gained via the methods and additional data used to develop the new infraction prediction tools.

The resulting models identified risk scores for each offender within a given infraction type. A scoring guide is provided, identifying risk points associated with each tool's items and responses. Offenders are to be

[1] Infraction outcomes are defined as Violent (any 1), Serious (any class 1 non-violent or 5+ class 2), and Non-serious (10+ class 3).

scored on each item and their scores summed. The summary score for each of the three infraction models is designed to place them into one of four categories – High Violent, High Serious, Moderate, and Low (see Figure 2).

In an effort to improve the classification system, the new tool is designed to inform and support classification staff efforts. Based on themes identified through the process evaluation, staff had indicated several issues that impacted the utility of Haryman tools' results. In particular, the scored classification designation is often overridden as a result of NDCS or offender needs (i.e. bed space availability and programming). Therefore, staff indicated a need for the ability to move offenders to custody designations based on rationales that are not solely based on security.

While infraction risk models were then developed. To ensure that the newly created tools were functional, provide face validity, and, in turn, gain user trust a cross-sectional group of NDCS staff were selected to review the tool. On July 26th the team of Subject Matter Experts (SMEs) gathered to review the tool and assess its functionality. SMEs were encouraged to provide feedback regarding the assessment items, usability, and overall design. Feasible adjustments to the tools were then completed and final models established.

Ultimately, the created tool provides categories that indicate an offender's infraction risk, instead of a one-to-one recommendation of custody designation. When used in conjunction with developed NDCS policy guidelines, the new classification schematic provides staff the flexibility to assign offenders to a lower/higher custody designation when agency or offender need requires. The new classification system also informs staff of an offender's likely infraction type and risk following a transfer to a new facility, providing the opportunity to differentiate supervision strategies once an offender is residing in their new facility. We feel this categorization system is a novel advancement of prior approaches. However, we also note that it represents a substantial variation from current practice and will require NDCS efforts around training and policy development to operate efficiently.

Recommendations

Following the initial development of the tools, we have outlined several recommendations. While some will be addressed in the next phases of the project, others identify long-term goals for greater prediction accuracy, more efficient uses of resources, and additional research needing completed.

1. *Create an implementation, training, and quality assurance plan*
2. *Continue improving the tool by adding items and collaborating with recent risk assessment efforts*
3. *Create efficient uses of assessment labor by identifying assessment redundancies*
4. *Create an inventory of interventions and forecast agency incarceration needs*
5. *Evaluate override factors and practices*

Next Steps

There are two remaining phases of the project. Manuals and training materials will need to be developed to adjust the current classification tools and identify any updated policies and new procedures. Materials developed will guide training of new staff as well as refreshers for current staff. Goals for booster training and other quality assurance guidelines will also be developed. All materials will be created in conjunction with NDCS SMEs.

Following the development of training materials, an implementation plan should be established. The implementation plan is recommended to include a timeline for a graduated roll out of the new tools, a pilot study to assess system impact and address any modification needs to scoring and cut point placement, a training schedule, and outline for quality assurance checks and future validation analyses.

INTRODUCTION

Statement of the Problem

Placement and movement of inmates within the Nebraska Department of Correctional Services (NDCS) incarceration facilities is an essential part of day-to-day operations. Inmates sentenced to a term of incarceration enter an NDCS diagnostic and evaluation facility, receiving a battery of assessments, recommendations, and ultimately an initial classification for appropriate placement in one of six custody designations. Following the initial placement, both routine and event driven reclassifications are provided. Over the last 45 years, the NDCS has made two substantial changes to its classification system. The first system was created and implemented in the 1970s. However, the primary issues of the system were that inmates were routinely over-classified and that substantial uses of overrides (approximately 40%) were indicated. It was therefore determined that this system required major improvements and lacked validation.

In 2005, Patricia Hardyman was contracted to update and modify the classification and reclassification system. Using statistical analyses of available data, a set of prediction models were created to score inmates on items that predicted future infraction behavior. Specifically, two infraction types were examined – any and violent infractions. A classic split sample validation procedure was used and produced four outcome models: male initial classification, male reclassification, female initial classification, and female reclassification. Both classification and reclassification tools are scored using a variety of risk factors intended to predict both violent and non-violent infraction behaviors while incarcerated (Hardyman, 2005). The risk factor item responses are summarized and the resulting score places males and females within a set of four ranges, indicating recommendations of: Community Custody (A or B), Minimum (A or B), Medium, and Maximum security. These scored custody designations may then be *overridden* based on 22 mandatory and discretionary override categories, assigning inmates to a security level not indicated by the scored value of the tool.

Although substantial improvements were made with Hardyman's updates, several methodological issues were identified. Specifically, override rates were still observed to be relatively high, where current estimates identify that roughly 40 percent of inmates are not classified at the level at which they score. Second, staff members are not trained on the tools intent and many were unaware that the intent of classification tool was to predict infractions. Furthermore, an updated validation had not yet been completed to ensure that the tool was operating as anticipated following implementation.

In addition, methodological issues related to tool development prevented the Hardyman tools from achieving a high rate of accuracy. Specifically, the number of predictors considered and the size of the development samples were relatively small for the purposes of creating an assessment. Also, the specificity of outcomes was not integrated into the prediction and classification of the tools created. In particular, models predicting infractions were not calibrated to the duration of time between initial and reassessment and the distinction of infraction types (i.e., violent, serious and any infractions) were not utilized when determining scoring categories. Furthermore, items were not weighted optimally to reflect the NDCS population. Finally, several methodological improvements in risk assessment development, not available at the time of Hardyman's tools' creation, are now feasible and can be incorporated into new tools' creation.

In conjunction with these tool related issues, a recent question was raised regarding prison overcrowding. A Justice Reinvestment report (2015) indicated that NDCS institutions possessed a substantial overcrowding issue, operating at 149% capacity. Additional concerns were raised regarding the frequency of unsupervised releases ("jam-outs"), insufficient provision of Evidenced-Based Programming (EBPs), and the potentially extensive costs of additional facilities needed to house incarcerated populations.

In response to these concerns, the NDCS identified the availability and flexibility of inmate assignment as a potential source of reducing issues surrounding overcrowding. In particular, scoring and override procedures of the current classification and reclassification instruments may provide an opportunity to streamline transfer processes. Specifically, the NDCS was interested in the following:

- An assessment of the validity of the current classification and reclassification tools, evaluating their ability to predict infraction behaviors.
- The construction and validation of new or modified classification tools with improved application usage and predictive accuracy.
- Provide recommendations for mechanisms and practices that would improve future classification efforts.
- Develop materials to be used by NDCS staff to train and assist in the adherence of best practices that surround classification scoring, data collection, and application of inmate risk scores.

Project Outline

The NDCS contracted with Zachary Hamilton, Ph.D. to complete the stated four deliverables. To provide a comprehensive assessment of the current tools, updated models, recommendations, and training materials, a multi-phased project was completed. Described in greater detail in the full report, the project parts consisted of the following.

1. **Process Evaluation** – This project was completed over several days of field research. First, classification and reclassification procedures were observed. Classification and diagnostic staff were interviewed. Inmate scoring, recommendation, and scoring procedures were also observed. In addition, focus groups consisting of classification staff, research, and administrative personnel were conducted to assess key advantages, disadvantages, and potential areas for improvement of the current classification system and processes. Finally, a document review of classification and reclassification procedures, prior research, and current data collection methods were completed. The culmination of findings outlined potential issues impacting accuracy and usage of scoring from both users as well as methodological limitations of current classification and reclassification model construction.

2. **Review of Quantitative Data and Analysis Plan** - Following the process evaluation, the availability of data needed for project deliverables was completed. This task consisted of an examination of current and potentially new items for consideration. Based on the evaluation of additional risk factors for potential inclusion in updated risk tools, a study sample was proposed. Ultimately, the sample consisted of 9,072 male and 1,582 female initial classification offenders incarcerated and supervised by the NDCS during the study period of August, 1991 and June, 2015. Due to the repeated use of reclassifications, an offender may receive more than one assessment. Thus, our reclassification sample consisted of 35,098 male assessments and 2,449 female assessments

   A total of 66 items were considered for the prediction of three outcomes, namely: Violent, Serious, and Non-Serious infractions. All outcomes were measured dichotomously (infraction/no infraction). For the majority of offenders, reclassifications are completed every six months. Therefore, a six-month follow-up interval was used to examine the specific risk of infraction for the duration leading up to the next reclassification assessment. However, some offenders were incarcerated for less than six months or did not possess full six-month duration until a subsequent reclassification. To account for these variations in incarceration exposure times, time-to-event modeling was utilized. This allowed for the joint assessment of both infraction occurrence and duration of time until the infraction was observed.

3. **Model Development** – Based on findings outlined in Phase II, infraction prediction models were constructed. First, the current classification and reclassification models were assessed for their ability

to predict infraction behavior. Next, feature selection procedures were completed, selecting items that improved (model fit) prediction of the three infractions outcomes. These models were created in an effort to improve upon, and replace, the current classification and reclassification models.

Additional models were created, examining the *reclassification* as the unit of analysis. Results from new models were created with the intention of providing staff and users added information. Specifically, these models allow staff to more accurately determine placement decisions based on offender-specific characteristics/behaviors since admission and the interactions of said characteristics/behaviors within a given custody level. The results of these models detail the likelihood of an individual inmate infracting in the six months following a reclassification assessment. All models were assessed via methodologically sophisticated validation procedures utilized and known to meet industry standards.

Comparisons were also made between the newly created models and those currently in use (e.g. the Hardyman models). Area Under the Curve (AUC) estimates, an industry standard for assessing prediction model accuracy/validity, were completed for the current (Hardyman) as well as all new male and female classification and reclassification models. Preliminary cut points were created that outlined the classification categories

4. **Subject Matter Expert (SME) Modifications** – Following model construction, preliminary findings were provided to NDCS Subject Matter Experts (SMEs). These SMEs were asked to provide feedback regarding the updated models and strategize mechanisms for adoption and implementation. Specifically, SMEs were asked to simulate classification with current cases and identify problems and/or adjustments to items, scoring, override usage and language. Feasible modifications were adapted as part of the new instruments' construction and validation efforts were completed for the modified tools.

5. **Conclusions and Use Recommendations** – Concluding our classification model efforts, recommendations are provided with regard to maintaining scoring fidelity via quality assurance (QA) and training boosters. Additional items and responses are also recommended to potentially enhance the tool going forward. The timing and use of routine assessments of interrater reliability and predictive validity are also suggested.

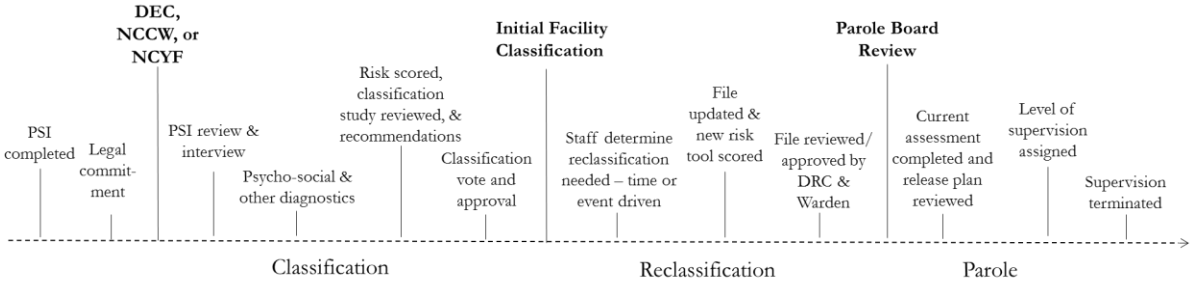
I – Process Evaluation

To provide a comprehensive review of the current classification tools' creation and usage, a process evaluation was conducted. This evaluation consisted of a thorough document review, classification and reclassification observations, classification and diagnostic staff were interviewed; inmate scoring, recommendation, and scoring procedures were also observed. In addition, informal focus groups were conducted, consisting of: classification staff, research, and administrative personnel to assess key advantages, disadvantages, and potential areas for improvement of the current classification system and processes. The current section provides a detailed description of findings provided via process evaluation efforts.

Synthesis of document review and observations

Extensive efforts were made to assemble and synthesize NDCS documents relevant to classification project goals. First, current classification policies, procedures, and forms were reviewed. This process provided information on facilities, resources, as well as common and uncommon methods in which inmates are housed and processed through the incarceration system. A synthesis of this review lead to the creation of the Classification Process Flow Chart, which is illustrated in Figure 1 and further described below.

Figure 1. NDCS Classification and Parole Process



Inmates sentenced to a term of imprisonment (legal commitment) are commonly held in a community facility (e.g. county jail) until sentenced and transferred to a diagnostic and evaluation center. Prior to their transfer, inmates are to receive a pre-sentencing investigation report (PSI), completed by a probation officer. The PSI is created to provide users with an inmate's criminal history, family, residential, and all other pertinent information to be used as part of their supervision classification, supervision and treatment planning. Inmates are then transferred to a diagnostic and evaluation facilities of which three exist in the state – the Diagnostic and Evaluation Center (DEC) for adult males, the Nebraska Correctional Center for Women (NCCW) and the Nebraska Correctional Youth Facility (NCYF) for youthful males adjudicated in adult courts.

Once they arrive, inmates are considered maximum custody until their classification has been approved and finalized. Classification staff members then examine the PSI when available. When the PSI is not available, staff must investigate the inmate's criminal and other pertinent information via available resources and data-bases. A *classification study* is begun by staff, which was described as the "Cliff Notes" of the full classification process. After the inmate's PSI information is entered, the staff members interview the inmate to verify the information, identify additional needs, and a Prison Rape Elimination Act (PREA) assessment is also completed. Medical and mental health staff members then meet with the inmate and complete a psycho-social interview (within 14 days) to assess additional issues and needs. If the inmate has a history of sexual or violent behavior, further assessments are completed by specialized classification teams (CSORT and CVORT). These additional assessments may be conducted using specialized instruments (i.e. PCL-R, Static-99, HCR-20, and others). These additional assessments are completed to provide more detail, informing the initial classification and future programming needs. The Classification Study is then completed,

which includes recommendations of programming, security restrictions, written comments from the inmate, and a score on the risk factor instrument.

The classification committee then meets to review the relevant information, discuss available alternatives and vote to determine the committee's final recommendation. At this stage the committee will assess the risk factor score/classification recommendation. If the inmate has policy mandated exclusions or additional issues to consider (beyond the risk factor score) a list of 22 mandatory and discretionary override rationales may be used to place the inmate in a custody level not identified by the risk factor scoring[2]. The inmate is then informed of the classification and transferred to an institution at their assigned custody level when available. If a facility at their assigned custody level is not available, the inmate waits at their current facility until placement in the recommended facility or an alternative facility (i.e. county jail) becomes available.

Dependent on their Tentative Release Date (TRD), the inmate is reclassified by staff at their current facility. Inmates with TRDs of more than three years are reassessed every 12 months while those with less time are reassessed every six months. Reclassifications are also completed when a serious infraction is committed or any other event necessitating an immediate reassessment of custody level. The assessment is used to determine if the inmate is eligible for transfer but may not necessarily change their custody level. Reclassification staff will gather information on the inmate since their last classification or reclassification. This will include but may not be limited to: infraction behavior, segregation and other housing restrictions (e.g., administrative segregation, intensive management, and protective custody), programming completed and needed. A reclassification specific risk scoring tool is used, which factors in program performance, infractions, and involvement with alcohol and drugs during their current incarceration. A Reclassification Narrative is also completed providing greater detail of the inmate's security and programming needs. It should be noted that parolees returning on a revocation are also provided a reclassification if their duration in the community was more than 12 months. The inmate's file is then reviewed by the Director's Review Committee (DRC), ultimately making a reclassification recommendation to the Warden. During this process, the inmate's score is reviewed and if needed, mandatory or discretionary overrides are indicated to adjust the final custody level. The reclassification process is repeated until the inmate's time is completed ("jam out") or they receive parole.

Classification Risk Scoring Tools

Next, classification and reclassification scoring tools and, where available, development materials were reviewed and evaluated. A total of four instrument versions were provided. Overall, several differences were identified among the four instrument versions reviewed. Common among all instruments were the following scoring items: severity of prior offenses, age, and escape history. Item variations identified across tool versions included: severity of current offense, number of prior convictions, past institutional behavior, age at first conviction, stability factors, involvement with drugs and alcohols, frequency of disciplinary infractions, severity of disciplinary infractions, performance in recommended work and/or programming, and projected length of incarceration. Mandatory overrides consisted of nine items, which included: death penalty case, misdemeanor or felony detainer and the sanction is fine/cost only, ICE detainer, detainer for low severity offense, detainer for moderate, high, or high severity offense, TRD greater than eight years, TRD greater than five years, TRD greater than three years, and administrative confinement/intensive management. Each tool also contained 13 discretionary override items, including: multiple failures to appear in last five years, pending investigation, program participation, NDCS need, medical condition, protective custody issues, mental health condition, non-compliance with program rules, central monitoring/separation issues, active

[2] Preliminary findings provided by the NDCS indicate that overrides occur for roughly 40% of the classification decisions.

participation in Security Threat Group (STG), detainer not accounted for by mandatory override, recent escape not accounted for by risk factors, and "other".

The Hardyman Report

The most extensive documentation was provided for the current instrument, contained within a report created by Dr. Patricia Hardyman (2006). Overall, the feature selection and development efforts are described in sufficient detail and well documented. Briefly, in 2004 and 2005 Dr. Hardyman and the NDCS convened a Steering Committee and reviewed the common issues and potential areas of improvement. A key issue identified was the Department's rate of overrides, noted at that time at 40 percent of all inmate classifications and reclassifications completed. Key development efforts were made to create classification instruments that were: accurate, reliable, appropriate for different prisoner populations, objective, simple to use, and efficient.

Hardyman provided a detailed analysis plan, indicating the steps and processes of creating the instrument scoring. First, data elements and resources were identified. While a majority of items used to score the inmates was provided and routinely entered in the management information systems (MIS), many items were only available in paper format and were manually coded. The samples drawn consisted of inmates incarcerated in an NDCS facility within the 2003 fiscal year (July, 2002 through June, 2003). A total of 210 female inmates and 1,677 male inmates were identified to form the construction sample. From this moderately sized purposive sample, a random sample of male offenders was drawn. This reduced the sample size of male offenders to 321 subjects. The female sample was relatively small from the outset and thus, the total female population was utilized. This construction sample was used to select infraction prediction items (feature selection). A second, validation, sample was also drawn. The intent of this sample was to test the created classification instrument. Again, a random sample of male offenders (n=358) was drawn and a purposive sample of female offenders was collected from the 2004 fiscal year (n=194)

Univariate assessments of all potential classification items to be used for scoring were assessed, and item response frequencies were further broken down by male and female sub-samples. A bivariate assessment of each item's (Pearson) correlation with violent and non-violent infraction misconduct reports was assessed and significant associations were noted. Items of significant and substantive importance were then entered into logistic regression models predicting violent and non-violent infractions for both male and female inmates in the 12 months following the classification/reclassification conducted. This resulted in a total of 4 regression models. Finally, cut point scores were established to determine infraction scoring ranges for community, minimum, medium and maximum custody level categories. Tests were computed from analysis of variance (ANOVA) models using validation samples to determine significant infraction prediction for inmates placed in the four scored custody categories.

Limitations of Hardyman Classification Tools

While substantial efforts were made to improve the predictive accuracy of the classification prediction scoring there are many notable limitations with the construction of the current classification tools. In addition, several recent advancements in prediction modeling can now be used to improve the classification system.

- ***Development sample size*** – Prior research has indicated sufficient sample sizes needed for prediction models of various populations (Styerberg et al., 2001). When the infraction outcome predicted is dichotomous (yes/no), a general rule is that a total of ten infractions is needed for each predictor variable. The limited sample size gathered for the study likely contributed to the small number of predictors utilized as part of the Hardyman models. Furthermore, the stability of prediction item estimates and validity of the models are also questionable with small validation

sample sizes. Unfortunately, a moderate sample size was obtained for male offenders; however, a random sample was used, which reduced the size of both the construction and validations samples. If a purposive sample was instead used, the issues that surround the randomly selected male sample would have been reduced.

- ***Prediction item specificity*** – The Hardyman models make use of a limited number of criminal and institutional history items. The limited number of items utilized created a specificity that would assist with prediction. For example, the severity of current and prior offenses and the number of prior offenses could be further broken down, creating separate items for several common event types, such as: assault, domestic violence, property, drug, sex, weapons, as well as distinctions for misdemeanors and felonies. Providing greater specificity improves the granular detail of infractions prediction and also allows for advancements regarding specified outcome prediction. This is but one example of how added item specificity will likely improve infraction model prediction. Other sources of item additions and specificity were also explored.

- ***Model weighting*** – Unweighted scoring systems use point systems that make no attempt to alter response scoring for item prediction strength. Said models simply add a point as response categories increase in severity. For instance, "current age" has five response categories and for each category increase, the inmate is given an additional point. By contrast, analytically weighted point systems make use of the statistical model estimates to adjust this weighting structure. These adjustments will improve the prediction accuracy of the scoring system, giving greater point values to items that are more influential for infraction prediction (Austin, Coleman, Peyton, & Johnson, 2003; Baird, 2009; Barnoski & Aos, 2003; Einhorn & Hogarth, 1975; Hamilton et al., 2016; Silver, Smith, & Banks, 2000). While it is clear that the Hardyman models adjusted the weights of some classification and reclassification responses, it is unclear how or why these weights were selected or if they are based on empirical evidence.

- ***Feature selection procedures*** – The feature selection process typically consists of gathering a large pool of potential scoring items and removing those that lack statistical or predictive performance. The Hardyman models used bivariate correlations to select features, or items, that were significant and substantial predictors of infractions. Reviewing the items included in the classification tools, there appears to be inconsistent logic as to why particular items are included in a given male or female model. That is, it appears as though some items (i.e. Time Remaining to Serve) were eliminated from the male classification tools due to a lack of bivariate significance while some female items are included despite non-significant findings. Furthermore, it has been shown that more accurate models are created when feature selection procedures use a multivariate approach, such as regression (Hamilton et al., 2016). By considering all potential items simultaneously, duplicative and unimportant measures, sometimes termed "noise items", are removed. In addition, multivariate models provide item specific weight estimates that can improve the accuracy of the instrument scoring structure.

- ***Inconsistent classification vs. reclassification item usage*** – An issue that is also reflected in staff comments was the lack of consistency in items scored as part of the classification and reclassification tools. The process of classification and reclassification differ, where reclassification has the benefit of using items collected since the inmate's admission date, such as: program participation, involvement with alcohol and drugs, infraction frequency, and infraction severity. However, items that are used in the initial classification are inexplicably removed from the reclassification tool, such as: stability factors, severity of current offense, escape history, number of prior convictions, and severity of prior

convictions. These inconsistencies between tools can result in dramatic scoring changes. In addition, items relating to infraction behaviors within the reclassification tool may be resulting in over-classification, as inmates in minimum and community custody facilities will often score out as medium or maximum following a Class II or III violation. Many items used in the initial classification may also be scored as part of reclassification assessments to eliminate inconsistencies and improve the accuracy and weighting of inmate characteristic/behavior items.

- ***Infraction outcome specificity*** – While there is no reference in the classification forms completed by NDCS staff, the Hardyman models were based on the prediction of the number of infraction misconduct reports (MRs) within 12 months of the classification/reclassification. As mentioned, bivariate correlations were completed as a part of feature selection, where each item considered for tool inclusion was assessed for the strength of association with violent and non-violent misconduct reports. However, the operational definitions of these two outcomes are inconsistent with the NDCS's three infraction classes. Specifically, Hardyman defined Violent MRs as a mix of Class I and II infractions, some of which are associated with violence but not a violent act in itself. In addition, the prediction of the number of any MRs is also questionable as there is no distinction between the three infraction class types. Furthermore, because the number of infractions is used, rather than a dichotomous measure of event occurrence, probabilities of infraction associated with each inmate score are not available and would likely provide useful information for classification staff and administrators. Ultimately, although two types of infraction MRs were used as part of feature selection, the tools do not provide separate scores predicting an inmate's risk of *violent* versus *any* infraction MR; thus, the inmate's computed score is simply *associated* with either or both MR outcome types.

- ***Less efficient validation techniques*** – Hardyman makes use of four Analysis of Variance tests (ANOVAs) to complete the validation of the tools, examining violent and any MRs for males and females. These four ANOVAs are repeated for the initial classification and again for the reclassification tool, for a total of eight models. ANOVAs are used to examine if categories of custody scores predict the number of MRs. There are several issues with this method of determining predictive validity. First, MRs are not normally distributed and thus violate a key ANOVA test assumption. Second, significance tests of risk categories are not commonly used as a validation procedure, as they are base-rate sensitive, where the industry standard method is to evaluate the offenders' raw scores using a Receiver Operating Characteristic (ROC) analysis, which produces an industry standard statistic – the Area Under the Curve (AUC). Lastly, the classic split sample method was used by Hardyman. More contemporary risk assessment model developments make use of more efficient k-fold methods and have been shown to provide more accurate estimates of predictive performance.

- ***Inability to adjust inmate scores for facility risk interactions*** – One of the key issues surrounding any classification system is the notion that the inmate changes and is changed by incarceration. Two well-known theories attempting to explain infraction behavior are *importation* and *deprivation*. Importation theory suggests that inmates bring their criminal attitudes from the community inside the facility, where the same offenders that are high risk for recidivism on the outside are high risk for infractions while incarcerated (Irwin & Cressey, 1962). The majority of the items collected as part of the NDCS tools are reflective of importation ideals. Deprivation theory suggests that the facility, its physical characteristics, community of offenders, availability of resources, and staff-to-inmate ratio also influence an inmate's likelihood of misconduct (Sykes, 1958). A key issue impacting the Hardyman created tools, and all others for that matter, is that scoring and

infraction prediction do not estimate or adjust inmate scores to account for the interaction of the facility they are to potentially being transferred. The acknowledgement of this inmate-facility interaction suggests that a given inmate may be more/less likely to infract if transferred to a facility with a lower/high security level or may have equal chance of infracting regardless of the custody level they reside.

Overall, care was taken by Hardyman to provide an infraction prediction tool that improved upon the prior tools using statistical modeling efforts. To the credit of these efforts, while much research has been published describing the underlying purposes and methodological advances of recidivism prediction instruments, little research has detailed the need, limitations, and development methods of correctional classification instruments. Furthermore, the method used by Hardyman, at the time, was likely considered sufficient and a progressive step beyond the previous instruments used. With that said, eight limitations have been noted as areas for improvement. While some of the alterations to be outlined may demonstrate greater improvement than others, the ultimate goal was to create a tool that, when implemented, will have greater utility for staff and administrators by providing a more accurate assessment of the inmate population.

Synthesis of Interviews and Focus Groups

Interviews and focus groups were completed to assess staff and administration's preferred uses of the current tools and the limitations. One-on-one interviews were conducted with key administrators, those involved in managing classification staff, and those involved in classification committees and final decision making. Staff members completing initial classification and reclassification assessments were also interviewed. In addition, psych associate staff were interviewed to assess processes for programming recommendations, specialized management considerations, and override recommendations. A total of three (informal) focus groups were also facilitated with classification and reclassification staff and administrators.

Classification scoring as a suggestion

Interviews and focus groups identified many common themes. Most described frustrations with the current scoring of the tools. Generally, staff indicated that the score computed by the Hardyman tools were not grounded in users' knowledge of the interpretation of the score, or what the score was intended to predict. While more detailed issues were described regarding the score calculation, the consensus was that the score itself seems somewhat arbitrary. With 22 mandatory and discretionary overrides available, the score produced by the Hardyman tools could be described best as a *suggestion* in which most classification staff were skeptical when offender scores did not match their internal calculation of risk.

A "bed-space driven process"

While referred to as a classification process, the ultimate goal is inmate movement and facility transfers, placing inmates where there is space available that is simultaneously the most appropriate yet least restrictive. A primary complication of the classification scoring is its utility to meet the transfer needs of the NDCS. Classification and reclassification was often described as a "bed-driven" process. Overcrowding could be observed in the initial Diagnostic and Evaluation Center (DEC), as inmates awaiting transfer may be required to sleep on cots and county jail contracts were used as a temporary stopgap for additional overcrowding. Those involved with the reclassification processes also described the pressure felt from the Ombudsman and DEC to reclassify and transfer inmates currently at medium and maximum security to minimum and community and note that lack of available beds creates difficulties when inmates are suitable for classification at a lower custody level. In an effort to create a smoother transfer to community correction, inmates approaching parole dates are often transferred from medium and maximum facilities straight to community corrections facilities. Programming availability and other facility restrictions sometimes required demotion from community or minimum to medium custody in order to receive needed interventions.

Discretionary overrides are commonly used to negate the reclassification score in order to meet these inmate movement demands.

Lack of specialized options and distinctions

The Lincoln Correctional Center (LCC) houses medium and maximum custody inmates. Here specialized housing can be provided for sex offenders and those with mental health issues. Apart from these options, few specialized housing options are available to provide graduated options or dedicated units for intervention purposes. Furthermore, with a few notable exceptions, custody distinctions within facilities provide no appreciable changes in restrictions/liberties. For example, an inmate can start out maximum and be reclassified to medium and remain in the same cell when beginning their next day in medium custody. Essentially, maximum and medium custody are the same custody level and similar non-distinctions may be observed in other facilities. Users suggested that more gradations in custody distinctions would allow inmates to "earn" their way to lower levels.

Mandatory Overrides

Another common issue with classification scoring are mandatory overrides. In particular, severity of offense, detainer status, and Tentative Release Date (TRD) seem to be a key area where over-classification may be occurring. As the intent of classification is infraction prediction, mandatory override selections are used to restrict transfers to lower custody levels due to important factors not considered as part of classification scoring. Classification staff indicated several issues with the current scoring procedures and the use of mandatory and discretionary overrides. With regard to overrides, the mandatory overrides for offense severity and TRD are not universally applicable and impact specific types of offenders. Staff commented that while sex offenders and long-term offenders often pose a lower security risk, mandatory overrides prevent their promotion to minimum security. In contrast, parole violators are often mandated to minimum/community custody due to TRD restrictions. While users understood the importance of these overrides to be used as safeguards, many questioned why these items were not used as scoring factors in either classification or reclassification. On several occasions groups discussed situations in which inmates convicted of serious offenses, those with minor detainers, and those with long TRDs would me of only a minor risk for misconduct.

Scoring

With regard to scoring items, several issues were indicated by SMEs. Often the severity of the inmate's misconduct is not consistently considered. A minor violation may be viewed as a demotion infraction for a community security facility but would be considered minor in a medium custody level. Furthermore, variations in abscond severity are not considered in reclassification scoring.

The inconsistencies between items used at initial versus those used at reclassification were also discussed. For instance, confinement in the community and prior incarcerations are only considered at reclassification, which may inflate an offender's score. The weighting structure of offender age greatly impacts younger offenders and may result in over-classification. Furthermore, classification staff indicated that detainer holds for fines, child support, and prior failure to appear events, do not reflect infraction behavior.

SMEs also identified some additional measures and modifications that could help improve the accuracy of scoring procedures. In particular, several additional assessments are completed by staff that could be added to the list of potential scoring items. For instance, the PREA assessment items could be included in classification and reclassification decisions. NDCS also provides a substantial resource with regard to verifying security threat groups, which could improve prediction. Staff also noted that time served could be a valuable predictor for future models. Staff also noted there is general lack of dynamic items that would allow an offender to reduce their classification score. Finally, there is a need to consider the recency of prior

infraction behavior, particularly for those inmates that have a tendency to self-sabotage their successful release/promotion.

Classification and Transfer Processes

Diagnostic staff repeatedly identified the (not so) timely delivery of the Presentencing Investigation (PSI) as problematic. When the PSI is delivered with the offender at reception classification is smooth and quick, however, when it is not then other (more time consuming) data sources are needed to complete the classification forms. Initial classification often feels rushed and the accelerated speed may prevent the identification of offenders' programming needs. A "one-stop-shop" should be created to allow for the timely delivery of data that would allow for the accurate complete of the classification study.

Programming and Availability

Staff also noted that even when programming and other inmate needs are identified, there are many restrictions that prevent the provision of recidivism and infraction reducing interventions. For instance, protective custody is only offered in medium and maximum security facilities and offenders with medical issues are not eligible for transfer to the Work Ethic Camp (WEC). Those offenders residing in higher custody levels often have few programs (beyond GED) in which to participate. Furthermore, when treatment needs are identified, the programming is often delayed substantially due to the offender's TRD.

Currently there is no design or policy that identifies programming timing/sequencing, which is problematic for offenders with multiple needs. Staff also indicated that program participation is scored arbitrarily, particularly when no programming is needed or available within a given facility. Finally, for sex offenders and inmates that have received a demotion, program participation, and improved behavior does not effectively translate into a transfer or return to a lower custody level.

Section Summary

The current section was used to outline both the use and need for custody classification tools. While the Hardyman tools were identified to possess substantial merit, notable areas of improvement need to be considered. Through interviews and focus groups several suggestions for improvements were provided. Although implementing all outlined suggestions may not be feasible for the current project, the next section describes the methods used to make predictive and process improvements to instrument scoring through the construction of new classification tools. For those issues not addressed through tool updates, the final section will provide recommendations for future research and policy modifications. In the next section, the study design and data sources are described.

II - Review of Quantitative Data and Analysis Plan

To examine the current validity of the Hardyman tool and potentially improve prediction strength, a comprehensive review of the NDCS agency records were examined. The intent was to identify a large sampling frame from which to create infraction prediction models. While the sampling frame was intended to be purposive, or remain as large as possible, the availability of predictor measures placed a number of restrictions on offenders' eligibility and the specificity of predictors to be included in the developed models.

A primary restriction was identified with regard to the reliability of infractions data prior to 1990. In addition, currently collected data on prior convictions does not indicate distinctions in severity (misdemeanor vs. felony), however, this level of detail is available for offenders' Current offenses. Agency records on escapes were also found to be inconsistently collected and deemed unreliable. As an outcome, violent misconduct was found to be relatively rare for female offenders, making a specified violent infraction prediction difficult to model for female offenders. Finally, the NDCS agency records were primarily restricted to static, criminal history items. Dynamic, offender needs-based items are relatively absent within the current predictor item pool.

Despite these restrictions, several additional measures, beyond those used by Handyman, were identified for inclusion in the predictor item pool. In particular, measures such as: Verified Security Threat Group, Sentenced for a Felony, Current offense, Earliest Release Date, Detainer Issue, Prior NDCS Incarcerations, the 22 discretionary and mandatory overrides, and an additional 13 measures that provide a more specified breakdown of type and seriousness of prior offenses and infraction behavior.

Measures

With regard to outcomes, three were identified to be of use to classification staff – violent, serious and non-serious infractions[3]. Violent infractions were defined as 1 or more Class 1 violent misconduct events that occurred within six months of the assessment. Serious infractions were defined as any Class 1 non-violent misconducts *or* five or more Class 2 infractions within six months of the assessment. The Non-Serious Infraction outcome was defined as 10 or more Class 3 infractions within six months of the assessment.

The six-month follow-up period was specified for each outcome to further increase the feasible use of the created tool. Specifically, the NDCS initial classification occurs in the first weeks of incarceration. For most offenders, a reclassification is completed six months following the initial classification and then again at each six-month interval of incarceration. Since the tool is designed to help predict an offender's behavior for six months only, the duration of the follow-up period, tracking infraction behavior, was capped at 182 days (previously 365). For offenders without a full six months of incarceration following an initial or reclassification, the exact number of days was recorded to accurately assess exposure time for each model.

With regard to model predictors, a large pool of measures was considered. In particular, over 70 items were considered in the prediction of each of the three outcomes. Operational definitions for all study measures are provided in Appendix I.

Sampling Frame

A generally appreciated part of both the Hardyman and prior tool construction was the separation of classification and reclassification assessments and gender specific scoring. To remain consistent with these tool components, four samples were created. First, due to the previously stated issue regarding data collection on infraction behavior, eligible offenders were those that received an initial classification prior to August of 1991. To allow for a sufficient duration of follow-up, an offender must have been admitted to prison prior to June of 2015, with the final tracking of infractions terminating in December of 2015. Next, those offenders

[3] The operational definitions for each infraction type were created through an examination of quantitative feasibility/use as well as contributions from SMEs.

that received an initial classification were identified and their infractions were tracked for their first six months of incarceration following the classification date. As indicated, the exact number of days incarcerated was also tracked for those offenders incarcerated fewer than 182 days. This final initial classification data set was further split into a male (n=9,072) and a female (n=1,582) sample.

Using the same eligibility criteria and infraction follow-up tracking processes, those offenders receiving a reclassification were identified for a separate sample frame. Each assessment was identified as a unit of analysis. However, offenders that spend greater than 18 months incarcerated will notably possess two or more reclassifications. This amounted to 35,098 male and 2,449 female reclassification assessments. Additional measures were used to examine the interaction of offender characteristics and prison environments. Thus the facility the offender was residing at the time of the reclassification and the facility in which they resided following the reclassification was identified[4].

Analysis plan

First, univariate analyses were completed to examine items for their relative prevalence within the sample. Based on this assessment, several items were removed from consideration due to their low prevalence (<1%) in the male or female sample. Next bivariate analyses (not displayed here) were completed to identify the ability of each item to predict at least one of the three infraction types. Where non-linear response patterns were identified, modifications to categories (collapsing and disaggregating techniques) were used to improve predictive performance of multivariate modeling techniques.

A series of Cox regression models were used to select and weight predictors. Cox regression is termed a *time-to-event* analysis and was ultimately selected to maximize sample size, including those offenders with a follow-up interval that were less than 182 days. Ultimately this analysis determines how each predictor impacts infractions while adjusting for individual exposure times.

Due to the large pool of potential predictors to be selected, we utilized a customized stepwise procedure. This procedure selects the most predictive indicators, one at a time, until all selected predictors improve model fit and those not selected, do not. However, item selection procedures that are purely data-driven can be problematic. That is, items may predict in an unanticipated direction, causing an illogical scoring schematic (Wainer, 1976). To adjust for this potential result, each predictor item is initially examined for theoretical/logical directionality. Items that are notable protective factors (i.e., higher education achievement, employment, etc.) are reverse-coded. This results in a model in which all selected measures weight in a consistent (positive) direction. All model predictors are described in Table 1, where reverse coded measures are indicated with an "R" following the item label.

We ultimately prevented the inclusion of illogically weighted items via a software solution. Using the *R* programming language, a selection procedure was developed to prevent items possessing a negative logit value from being included. In addition, items were also selected based on model improvement identified via their ability to improve the AUC statistic, removing predictors that do not provide an incremental improvement to the model. As noted earlier, the AUC is considered the industry standard when creating prediction tools and identifies the joint assessment of each item's ability to improve prediction specificity and sensitivity. Using these two criteria – positive item logit and model AUC value improvement – this customized procedure extends modeling procedures developed previously (Hamilton et al., 2016).

Table 1. Sample Descriptives – Initial Classification

Predictors	**Female (n=1,582)**	**Male (n=9,072)**	**Total (N = 10,678)**
Race†			
White/Caucasian	63.0	57.5	58.3
Black/African American	17.4	24.0	23.0

[4] If the offender was not transferred following the assessment, this same facility was indicated as their residence following the reclassification.

Hispanic	8.0	13.5	12.7
Native American	7.3	3.8	4.4
Other	4.3	1.2	1.7
Security Threat Group - Verified	2.3	19.4	16.9
Sentenced for Felony	86.9	86.4	86.5
Current offense (not mutually exclusive)			
Violent	19.4	28.6	27.2
Property	35.0	25.4	26.9
Drug	39.8	37.0	37.4
Sex	3.9	9.7	8.8
Other	12.5	11.7	11.8
Age At Admission – R			
<18	0.2	1.0	0.9
18-19	3.8	5.7	5.4
20-29	37.1	41.0	40.5
30-39	29.9	25.6	26.3
40-49	22.5	17.6	18.4
50-59	5.6	7.5	7.2
60+	0.9	1.5	1.4
Age of First Conviction – R			
24+	30.5	13.9	16.4
19-23	28.3	22.6	23.5
15-18	30.3	41.4	39.7
<15	11.1	22.0	20.4
Highest Grade Completed – R			
Some College or More	23.3	27.3	26.7
High School Diploma or GED	23.0	24.6	24.4
11th Grade or Less	53.7	48.1	48.9
Full Time Employment/Child Care Prior – R	28.7	32.1	31.6
Earliest Release Date – R			
< 6 months	25.0	28.4	27.9
6-12 months	29.9	26.5	27.0
13-18 months	18.1	14.8	15.3
19-24 months	12.3	12.0	12.0
25-36 months	10.5	10.9	10.8
37+ months	3.8	6.4	6.0
Life	0.3	1.1	0.9
Detainer Issue	29.0	34.1	33.4
Prior NDCS Incarcerations			
0	78.8	65.8	67.7
1	14.5	20.0	19.2
2	4.2	8.5	7.8
3	1.4	3.1	2.8
4+	1.1	2.6	2.4
Seriousness of Current offense			
Misdemeanor	13.1	16.3	13.5
Felony 3 or 4	79.4	77.0	77.4
Felony 1 or 2	7.5	9.3	9.1
Prior Violent Offense			
0	70.8	34.1	39.6

1	18.2	25.0	24.0
2	6.3	15.6	14.2
3	2.6	9.3	8.3
4	0.9	5.5	4.8
5	0.5	3.0	2.6
6	0.2	1.6	1.4
7	0.2	1	0.8
8+	0.4	4.9	4.3
Prior Property Offense			
0	36.8	32.8	33.4
1	18.2	18.5	18.5
2	13.4	12.6	12.7
3	7.7	8.8	8.7
4	6.6	5.7	5.9
5	2.7	3.8	3.7
6	2.5	2.4	2.4
7	1.6	1.3	1.4
8+	10.5	13.9	13.4
Prior Drug Offense			
0	38.3	19.4	22.2
1	20.3	17.0	17.5
2	13.3	14.7	14.5
3	9.6	12.0	11.6
4	6.6	8.8	8.5
5	3.5	5.9	5.6
6	2.2	4.0	3.7
7	0.7	2.5	2.2
8+	5.6	15.8	14.3
Prior Sex Offense			
0	82.1	75.4	76.4
1	12.0	15.9	15.3
2	2.6	4.6	4.3
3	0.9	1.9	1.7
4	0.5	1.1	1.0
5+	1.9	1.2	1.3
Prior Weapons Offense			
0	95.4	79.3	81.7
1	3.8	14.8	13.1
2+	0.8	5.9	5.2
Prior Escape Offense			
0	96.2	93.3	93.7
1	3.5	5.6	5.3
2+	0.3	1.1	1.0
Prior Domestic Violence Offense			
0	95.6	81.2	83.4
1	3.9	11.9	10.7
2+	0.5	6.8	5.9
Prior Assault Offense			
0	71.8	39.2	44.0
1	16.7	24.4	23.2

2	6.3	13.5	12.5
3	2.1	8.2	7.3
4	1.5	5.2	4.7
5+	1.6	9.5	8.3
Prior Violent-Property Offense			
0	85.3	57.3	61.5
1	11.2	24.4	22.4
2	2.3	10.4	9.2
3	0.8	4.5	3.9
4	0.3	1.7	1.5
5+	0.1	1.8	1.6
Prior Any Offense			
<4	31.3	18.1	20.1
4-6	21.0	18.1	18.5
7-10	18.8	22.5	21.9
11-15	13.8	18.1	17.4
16+	15.1	23.3	22.1
Prior Any Infractions			
0	31.4	18.1	20.0
1-4	21.0	18.1	18.5
5-9	18.8	22.5	21.9
10-24	13.8	18.1	17.5
25+	15.0	23.3	22.1
Prior Any Class 3 Infractions			
0	86.3	79.8	80.8
1-10	9.5	8.1	8.4
11-25	2.7	4.6	4.3
26-50	1.0	3.7	3.3
51+	0.4	3.7	3.3
Prior Incarceration Violent Infractions			
0	99.4	95.3	95.9
1	0.4	2.8	2.5
2+	0.1	1.9	1.6
Prior Incarceration Serious Infractions			
0	94.1	87.6	88.6
1-5	5.8	9.3	8.8
6-10	0.1	1.8	1.6
11+	0.0	1.3	1.1
Prior Incarceration Serious Class 2 Infractions			
0	89.6	81.1	82.4
1-5	7.1	7.4	7.4
6-15	2.3	5.3	4.8
16+	0.9	6.2	5.4
Approved Custody Level			
Max	4.5	6.6	6.2
Medium	22.4	15.8	17.4
Minimum A	35.2	41.0	39.6
Minimum B	11.4	2.3	4.4
Community A	26.5	17.9	19.8
Community B	0.0	16.3	12.6

Discretionary Overrides			
D1 – Multiple (4+) FTAs	0.1	0.2	0.1
D2 – Pending Investigation	0.3	0.2	0.2
D3 – Program Participation	26.9	18.6	19.9
D4 – NDCS Need	4.0	11.8	10.7
D5 – Medical Conditions	3.2	0.4	0.8
D6 – Protective Custody	1.5	4.3	3.9
D7 – Mental Health Concern	0.6	0.5	0.5
D8 – Non-Compliance with Program Rules	1.8	1.2	1.3
D9 – Central Monitoring/Separation Issues	0.2	0.4	0.4
D10 – Security Threat Group	0.1	0.4	0.4
D11 – Active Detainer/Pending Charges	0.3	1.4	1.2
D12 – Escape Threat	0.2	0.2	0.2
D13 - Other	4.7	9.4	8.7
Mandatory Overrides			
M1 – Death Penalty Case	0.1	<0.1	<0.1
M2 – Detainer with Fines Only	0.9	0.7	0.7
M3 – ICE Detainer	1.3	5.0	4.5
M4 – Low Severity Detainer	7.6	4.7	5.1
M5 – Moderate/High Severity Detainer	17.3	8.3	9.6
M6 – Tentative Release Date 8+ Years	0.6	1.3	1.2
M7 – Tentative Release Date 5+ Years	0.7	0.8	0.8
M8 – Tentative Release Date 3+ Years	1.9	2.4	2.3
M9 – *Administrative Confinement*	1.2	3.0	2.8
Hardyman Items			
Overall Score	31.8 (0.1)	25.1 (0.1)	NA
1 – Severity of current offense			
Highest	12.9	21.2	20.0
High	4.3	13.3	11.9
Moderate	52.1	41.4	43.0
Low	30.7	24.1	25.1
2 – Number of Prior Convictions			
5+	51.3	57.9	56.9
1-4	33.5	30.4	30.8
0	15.2	11.7	12.2
3 – Severity of Prior Convictions			
Highest/High	24.3	39.5	37.2
Moderate	48.0	31.7	34.1
Low/None	27.6	28.9	28.7
4 – Escape History			
Secure w/in 5 Years	0.8	0.5	0.5
Secure 5-12 Years	0.7	0.4	0.5
Non-secure w/in 3 Years	2.1	1.8	1.9
Non-secure 3-7 Years	1.4	1.9	1.8
None	95.1	95.4	95.3
5 – Past Institutional Behavior			
Class I – Violent 10 Years	0.6	1.6	1.4
Class II – Violent 2 Years	1.0	2.5	2.3
Class I/II – Nonviolent 2 Years	2.9	4.1	3.9

None	95.4	91.9	92.4
6 - Age at First Conviction – R			
<20	49.7	66.9	64.3
20-27	31.1	22.8	24.1
28-38	13.3	7.1	8.0
39+	6.0	3.2	3.6
7 – Current Age – R			
Female <25/Male <23	21.8	24.8	24.4
Female 25-43/Male 24-32	60.2	32.1	36.3
/Male 33-42	0.0	21.7	18.5
Female 44-50/43-55	12.7	18.3	17.5
Female 51+/Male 56+	5.3	3.0	3.4
8 – Stability Factors – R			
0 – Neither	32.1	20.4	22.2
1 - Ged/High School Diploma or Employed full time at arrest/Child care	58.7	44.3	46.4
1 - Ged/High School Diploma and Employed full time at arrest/Child care	9.2	35.3	31.4
9 – Projected Length of Incarceration (female only) – R			
73 months or more	0.9	--	--
48-72 months	2.2	--	--
19-47 months	38.7	--	--
18 months or less	58.2	--	--
Outcomes (6 months)			
Mean Exposure Days	170.1 (0.7)	170.3 (0.3)	170.7 (0.3)
Non-Serious Infractions Hazard Ratio	1.6	0.6	1.0
Violent Infraction Hazard Ratio	0.5	2.1	1.0
Serious Infraction Hazard Ratio	1.0	1.0	1.0
Mean Survival Non-Serious Infractions	166.9 (0.8)	171.8 (0.3)	171.1 (0.3)
Mean Survival Violent Infraction	176.8 (0.5)	173.4 (0.3)	173.9 (0.3)
Mean Survival Serious Non-Violent	163.4 (0.9)	163.1 (0.4)	163.1 (0.4)

R - indicates items that are reverse coded

†It should be noted that Race/Ethnicity are provided here for descriptive purposes but this item was not considered for inclusion as a predictor in any model.

A total of 12 classification models were completed. For the initial classification tool, a model for each of the three outcomes was completed for both males and females, representing 6 of the 12 models created. An additional six were created for reclassification. Regarding the reclassification sample, a model for each of the three outcomes was completed for both males and females. For the reclassification sample, following the selection of predictors, each model was also examined for potential interaction with the offenders' post-reclassification transfer facility. Sample descriptives for the reclassification sample are presented in Table 2.

Table 2. Reclassification Descriptives

Predictors	**Female (n= 2,449)**	**Male (n=35,098)**	**Total (N =37,098)**
Race			
White/Caucasian	63.3	55.5	56.1
Black/African American	16.8	26.0	25.4
Hispanic	8.0	13.0	12.7
Native American	8.2	4.4	4.7
Other	3.7	1.1	1.2

Security Threat Group - Verified	3.2	31.9	30.1
Sentenced for Felony	91.5	90.6	90.7
Current offense (not mutually exclusive)			
Violent	21.6	42.7	41.3
Property	34.1	25.3	25.9
Drug	39.6	25.1	26.1
Sex	4.9	13.8	13.2
Other	11.1	10.0	10.1
Age At Admission – R			
<18	0.0	2.6	2.5
18-19	4.0	8.3	8.0
20-29	39.6	44.2	43.9
30-39	29.2	23.8	24.1
40-49	22.1	14.6	15.1
50-59	4.8	5.4	5.3
60+	0.3	1.2	1.1
Age of First Conviction – R			
24+	26.2	10.2	11.3
19-23	29.6	18.9	19.6
15-18	33.8	41.0	40.5
<15	10.4	29.9	28.6
Highest Grade Completed – R			
Some College or More	29.2	25.3	25.6
High School Diploma or GED	25.8	25.7	25.7
11th Grade or Less	45.0	49.0	48.7
Full Time Employment/Child Care Prior – R	--	--	--
Earliest Release Date – R			
< 6 months	35.6	42.1	41.7
6-12 months	25.0	17.1	17.7
13-18 months	13.6	9.3	9.6
19-24 months	10.7	7.5	7.7
25-36 months	8.1	6.9	7.0
37+ months	7.0	6.1	6.1
Lifer	0.0	11.0	10.3
Detainer Issue	29.3	38.1	37.6
Prior NDCS Incarcerations			
0	75.5	62.9	63.7
1	14.7	21.6	21.2
2	6.8	9.0	8.8
3	1.8	4.0	3.9
4+	1.3	2.5	2.4
Seriousness of Current offense			
Misdemeanor	8.5	9.4	9.3
Felony 3 or 4	79.2	69.8	70.4
Felony 1 or 2	12.3	20.8	20.2
Prior Violent Offense			
0	68.5	28.2	30.8
1	20.5	25.4	25.1
2	7.0	15.1	14.6
3	2.2	10.9	10.3

4	0.8	7.0	6.6
5	0.2	4.2	3.9
6	0.0	2.2	2.0
7	0.1	1.3	1.2
8+	0.7	5.8	5.4
Prior Property Offense			
0	33.9	30.7	30.9
1	18.6	17.5	17.6
2	13.5	12.3	12.4
3	9.4	9.3	9.3
4	5.8	5.8	5.8
5	2.8	4.4	4.3
6	3.2	3.1	3.1
7	1.3	1.8	1.8
8+	11.5	15.1	14.9
Prior Drug Offense			
0	41.4	26.7	27.6
1	17.4	17.2	17.2
2	12.2	13.9	13.8
3	8.2	10.7	10.5
4	7.1	7.6	7.6
5	3.8	5.0	5.0
6	2.8	3.3	3.2
7	1.2	2.3	2.3
8+	5.8	13.3	12.8
Prior Sex Offense			
0	82.5	71.4	72.2
1	12.5	18.6	18.2
2	1.6	5.8	5.5
3	1.3	1.9	1.9
4	0.9	1.1	1.1
5+	1.1	1.1	1.1
Prior Weapons Offense			
0	94.9	75.0	76.3
1	4.3	16.6	15.8
2+	0.8	8.4	7.9
Prior Escape Offense			
0	97.7	91.1	91.6
1	2.2	7.0	6.7
2+	0.1	1.9	1.7
Prior Domestic Violence Offense			
0	97.4	85.1	85.9
1	2.6	9.6	9.1
2+	0.0	5.4	5.0
Prior Assault Offense			
0	71.4	36.6	38.8
1	16.7	25.8	25.2
2	6.2	13.2	12.7
3	2.8	8.7	8.4
4	1.2	5.3	5.0

5+	1.8	10.4	9.8
Prior Violent-Property Offense			
0	84.6	52.3	54.4
1	12.0	25.1	24.3
2	2.4	11.9	11.3
3	0.4	5.7	5.4
4	0.4	2.4	2.2
5+	0.2	2.6	2.4
Prior Any Offense			
<4	30.3	20.4	21.0
4-6	19.6	16.4	16.6
7-10	20.0	20.4	20.3
11-15	14.3	16.9	16.7
16+	15.8	26.0	25.4
Prior Incarceration Any Infractions			
0	30.3	20.4	21.0
1-4	19.6	16.4	16.6
5-9	20.0	20.4	20.3
10-24	14.3	16.9	16.7
25+	15.8	26.0	25.4
Prior Incarceration Violent Infractions			
0	99.5	91.9	92.4
1	0.4	4.0	3.8
2+	0.2	4.1	3.8
Prior Incarceration Serious Infractions			
0	91.6	83.1	83.7
1-5	8.2	10.4	10.3
6-10	0.2	3.4	3.2
11+	0.0	3.1	2.9
Any Infractions During Current Incarceration			
0-5	32.3	24.3	24.8
6-20	33.4	24.5	25.1
21-50	21.6	18.8	19.0
51-100	7.8	13.8	13.4
101+	4.9	18.7	17.8
Violent Infractions During Current Incarceration			
0	91.3	67.0	68.6
1-2	6.1	15.6	15.0
3-4	1.9	12.3	11.6
5+	0.7	5.1	4.8
Serious Infractions During Current Incarceration			
0	59.2	35.9	37.4
1-2	27.6	22.8	23.2
3-5	8.3	15.2	14.8
6-12	2.9	13.0	12.4
13-30	1.9	8.7	8.2
31+	0.0	4.3	4.0
Violent/Serious Infractions During Current Incarceration			
0	59.2	35.9	37.4
1-2	25.8	19.5	19.9

3-5	9.4	15.5	15.1
6-12	3.5	14.2	13.5
13-30	2.0	9.5	9.0
31+	0.0	5.4	5.0
Any Infractions 6 Months Prior to Reclassification			
0	27.1	29.0	28.9
1-4	26.9	26.2	26.2
5-11	23.6	23.4	23.4
12+	22.4	21.4	21.5
Violent Infractions 6 Months Prior to Reclassification			
0	96.6	91.3	91.7
1	3.2	7.3	7.1
2+	0.2	1.3	1.3
Serious Infractions 6 Months Prior to Reclassification			
0	78.4	68.2	68.9
1-2	18.6	23.2	22.9
3-5	2.7	6.9	6.7
6+	0.4	1.6	1.5
All Prior Any Infractions			
0-5	32.9	22.5	23.1
6-20	28.5	19.5	20.0
21-50	24.5	19.5	19.8
51-120	9.5	18.8	18.2
121+	4.7	19.8	18.8
All Prior Violent Infractions			
0	91.0	63.0	64.8
1-2	7.6	24.2	23.1
3-5	1.4	7.7	7.3
6+	0.0	5.1	4.8
All Prior Serious Infractions			
0	54.4	30.5	32.0
1-2	30.2	21.9	22.5
3-5	9.3	15.9	15.4
6-10	4.0	12.3	11.7
11-25	2.1	11.4	10.8
26-60	0.0	6.1	5.7
61+	0.0	2.0	1.8
Class 2 Serious Infractions Prior to Reclass			
0	84.6	77.4	77.9
1-5	11.4	7.7	8.0
5-15	3.1	5.6	5.5
16-40	0.7	4.8	4.6
41+	0.2	4.4	4.1
10+ Class 3 Infractions Prior to Reclass			
0	81.1	76.3	76.6
1-10	12.4	8.3	8.6
11-30	5.2	5.8	5.8
31-75	1.2	5.0	4.7
76+	0.2	4.5	4.3
Incarceration Duration to Reclass			

<6 months	33.4	23.1	23.7
<6 months to 1 year	26.4	19.1	19.5
1 to 2 years	21.0	20.4	20.4
2-5 years	14.3	19.9	19.6
5+ years	4.9	17.5	16.7
Approved Custody Level			
Max	14.8	14.7	14.7
Medium	25.7	22.1	22.4
Minimum A	9.5	1.2	1.8
Minimum B	25.8	22.5	22.8
Community A	16.6	24.6	24.0
Community B	7.7	14.9	14.4
Discretionary Overrides			
D1 – Multiple (4+) FTAs	0.0	0.1	0.1
D2 – Pending Investigation	0.4	0.3	0.3
D3 – Program Participation	40.3	30.3	31.0
D4 – NDCS Need	5.4	18.9	18.0
D5 – Medical Conditions	3.3	0.6	0.8
D6 – Protective Custody	2.2	10.7	10.1
D7 – Mental Health Concern	0.2	1.2	1.1
D8 – Non-Compliance with Program Rules	4.7	2.0	2.2
D9 – Central Monitoring/Separation Issues	0.1	0.5	0.5
D10 – Security Threat Group	0.0	1.0	0.9
D11 – Active Detainer/Pending Charges	0.2	1.2	1.1
D12 – Escape Threat	0.1	0.2	0.2
D13 – Other	14.7	15.6	15.6
Mandatory Overrides			
M1 – Death Penalty Case	0.2	0.3	0.3
M2 – Detainer with Fines Only	1.1	0.7	0.7
M3 – ICE Detainer	0.7	2.7	2.6
M4 – Low Severity Detainer	5.5	4.2	4.3
M5 – Moderate/High Severity Detainer	16.3	8.2	8.7
M6 – Tentative Release Date 8+ Years	0.7	10.8	10.1
M7 – Tentative Release Date 5+ Years	4.3	9.4	9.1
M8 – Tentative Release Date 3+ Years	8.0	8.2	8.2
M9 – *Administrative Confinement*	3.4	12.9	12.3
Hardyman Items			
Overall Score	28.0 (0.1)	26.5(0.1)	26.6 (0.1)
1 – Involvement with Drugs or Alcohol			
Two or More	2.5	2.9	2.8
One	7.4	8.9	8.7
None	90.2	88.3	88.4
2 – Escape history			
Secure within 5 years	0.5	0.9	0.9
Secure within 5-12 years	0.5	1.0	0.9
Non-secure within 3 years	5.0	4.1	4.2
Non-secure 3-7 Years	1.0	1.8	1.7
None	93.0	92.2	92.3
3 – Frequency of Disciplinary Infractions			

Two or more in last 12 months	22.4	26.6	26.2
One in the last 122 months	15.3	16.7	16.6
None in the last 6 months	4.5	6.1	6.0
None in the last 12 months	57.8	50.6	51.2
4 – Severity of Disciplinary Infractions			
Class I – Violent	1.4	5.1	4.8
Class I – Non-Violent	5.9	4.7	4.8
Class II – Violent	5.0	10.9	10.4
Class II – Non-Violent	29.5	27.4	27.6
Class III	34.9	25.9	26.7
None	23.3	26.0	25.8
5 – Conviction History			
Highest	20.8	48.4	46.1
High	13.3	18.3	17.9
Moderate	53.8	22.2	24.9
Low/None	12.1	11.1	11.2
6 – Current Age – R			
Female <25/Male <23	16.4	19.9	19.6
Female 25-43/Male 24-32	63.1	33.7	36.2
/Male 33-42	--	24.1	18.0
Female 44-50/43-55	15.3	18.2	4.2
Female 51+/Male 56+	5.1	4.1	
7 – Performance in Work/Programming – R			
Refused program/work or was terminated last 6 mon	25.3	19.9	20.4
Selective compliance/plan, waiting list, or working	30.3	47.9	46.4
Working and compliant with plan	44.4	32.2	33.2
Reclass Facility Location			
CCL	11.9	3.1	3.6
CCO	1.5	0.5	0.6
DEC	--	63.6	60.1
LCC	--	4.8	4.5
NCW	77.0	--	4.3
NYC	--	1.1	1.1
NSP	--	12.3	11.6
OCC	--	7.9	7.4
TSC	--	4.7	4.4
WEC	9.6	2.1	2.5
Facility Location Following Reclass			
CCL	42.0	16.1	17.8
CCO	10.5	5.9	6.2
DEC	0.5	4.7	4.4
LCC	--	7.7	7.2
NCW	38.6	--	2.5
NYC	--	2.3	2.1
NSP	--	26.2	24.5
OCC	--	18.1	16.9
TSC	--	16.2	15.7
WEC	--	2.2	2.6
Transferred to a New Facility Following Reclass	70.0	83.7	82.8

Outcomes (capped at 6 months)			
Mean Exposure Days	152.7 (0.3)	162.9 (0.9)	152.0 (0.3)
Non-Serious Infractions Hazard Ratios	1.2	0.8	1.0
Violent Infraction Hazard Ratios	0.9	1.1	1.0
Serious Infraction Hazard Ratios	1.0	1.0	1.0
Mean Survival Non-Serious Infractions	178.4 (0.4)	179.5 (0.1)	179.5 (0.1)
Mean Survival Violent Infraction	178.3 (0.5)	177.6 (0.1)	177.7 (0.1)
Mean Survival Serious Non-Violent	161.0 (1.0)	162.3 (0.3)	162.2 (0.3)

R - indicates items that are reverse coded

†It should be noted that Race/Ethnicity are provided here for descriptive purposes but this item was not considered for inclusion as a predictor in any model.

Validation

Assessing the predictive performance of each model was conducted using a validation technique referred to as K-fold cross-validation. Generally, there are two steps needed to validate a risk assessment instrument: training of the risk model based on a set of data and then testing the created models on a new set of data that the model has never *seen* before (to assess how well it makes correct predictions). Simpler methods that employ this technique often use a split-sample procedure, separating the data into two equal halves: one for training, the other for testing. The limitation with this method is that it does not use all of the data available for each of the two steps, only one half.

A method that resolves this limitation is 10-fold cross validation, which partitions the dataset into 10 equal parts at random. Nine of the parts are used for training the risk model, with the remaining part used for testing. This process is then replicated/repeated 10 times, with a different tenth of the data used for testing each time. The performance metrics of the predictions for each of the 10 subsets are then summarized to yield a single score. The performance metric used was the ROC curve and its associated AUC statistic. To briefly describe the meaning of the values calculated, an AUC of .5 would indicate model prediction relatively equal to that of chance, whereas an AUC of 1.0 would represent perfect predictive accuracy. Industry standard identifies four ranges/effect sizes of AUC values – negligible (<.56), small (0.56-0.63), moderate (0.64-0.70), and large (>0.71) (see Rice and Harris, 2005).

Model Comparison

Following model construction, the validation AUC statistics for all models were then compared. That is, AUC statistics for the newly constructed models were computed to that of the Hardyman initial classification and reclassification scales. Updated Hardyman scales were also constructed using the current outcome definitions and the more advanced modeling procedures described above. When comparing AUC statistics between models, greater values and category effect sizes identify improved prediction strength.

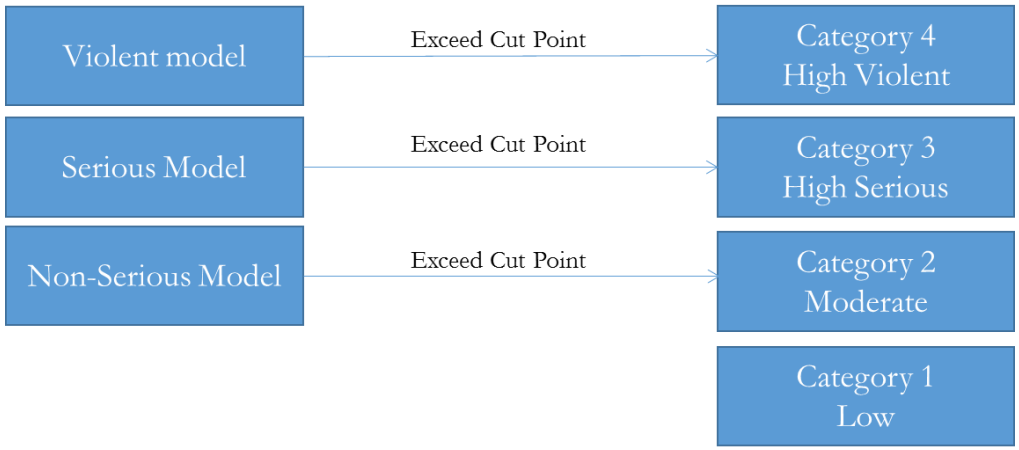
Risk Category Construction

Finally, model cut points were established, identifying high versus low risk offenders for each infraction type. High risk cut points for each model were established by identifying the odds of committing an infraction that was at least twice the hazard rate of the low risk group. Briefly, a hazard ratio is used in time-to-event modeling and indicates relative odds of the high risk group failing (infracting) as compared to those not identified to be high risk.

To combine the effects of the three models a hierarchical classification structure was then assembled, prioritizing high risk for violent, then serious, followed by any infractions. A depiction of this classification schematic is illustrated in Figure 2 below. Classification rules are described as follows. Offenders identified to be high risk in the Violent infraction model are placed into Category 4 – *High Violent*. Those not identified as Category 4 but were identified as High Risk within the Serious infraction model are identified as Category 3 – *High Serious*. Offenders not classified as high risk in the prior two outcome models were assessed with the Non-Serious infraction model and those exceeding the cut point are identified as Category 2 – *Moderate Risk*. Those offenders not classified in the previous models are identified as Category 1 – *Low Risk*. Because categories are assembled into a single hierarchical scale, an offender may fall into the high-risk category (e.g.,

Category 3) but may also be eligible for Category 2 placement. Based on the hierarchy of categories, offenders are placed in the higher prioritized category (i.e., Category 3). For each category, sample proportions as well as the hazard ratios provided.

Figure 2. Risk Category Creation Process



Section Summary

The current section outlined the study design, including the measures, sampling frame and analysis plan. Descriptive statistics for the male and female initial and reclassification samples were also provided. Finally, the hierarchical risk categorization was described. In the next section model results will be provided. This includes both initial and reclassification models and cut points. The models constructed are labeled as "preliminary", where SME input and model modifications will be used to guide final model construction.

III – Model Development

As indicated in the analysis plan, several infraction assessment models were evaluated and considered. First, the Hardyman tool was examined, identifying its summary score's ability to predict infraction behavior. Next, to provide a comparison of methods, the Hardyman items were included in multiple regression analyses. Finally, new models were constructed using the Hardyman items and additional, available measures routinely collected by the NDCS. Each assessment model was assessed for its ability to predict three infraction outcomes – violent, serious, and non-serious infractions. All models were analyzed using a Cox regression time-to-event modeling procedure. Each of the models was computed separately for male and female samples. Results from the initial classification models are presented first, followed by reclassification findings.

Initial Classification

Hardyman Scores and Models

First, using the Hardyman items, summary scores were calculated for each offender. These scores represent those calculated by NDCS assessment staff. The summary scores were then entered into each one of the survival models. AUC statistics for each model are presented in Table 3. To remind readers, industry standards identify four ranges/sizes of AUC effects – negligible (<.56), small (0.56-0.63), moderate (0.64-0.70), large (>0.71) (see Rice and Harris, 2005). When examining the AUC statistics from the Hardyman scores, five of the six models are identified to provide a small effect size with the sixth (Female Violent) identified as moderate. Using these industry standards as a reference point, the Hardyman scores do not provide optimal prediction strength for the three infraction outcomes.

Table 3. Hardyman Score and Survival Initial Classification Infraction AUC

Model	Violent	Serious	Non-Serious
Female	0.66	0.58	0.59
Male	0.62	0.58	0.61

As discussed previously, the Hardyman models were not developed with currently available methodological techniques. Furthermore, Hardyman did not construct models to predict the specific outcomes identified here. Without optimizing the prediction items to the outcomes, item weights do not achieve maximum performance and unneeded items may add noise to the prediction score calculated for each offender. To adjust for these methodological limitations, the Handyman items were entered separately into three multiple regression survival models (one for each outcome). Using the customized selection procedure (previously described), items that improved predictive performance were identified and weighted. The resulting model items, weights, and AUC statistics are provided in Table 4.

Table 4. Hardyman Models and Survival Initial Classification Infraction AUC

	Male			Female		
Item	**Violent**	**Serious**	**Non-Serious**	**Violent**	**Serious**	**Non-Serious**
1 – Severity of current offense	0.09	0.05	0.09	0.17	0.08	0.11
2 – Number of Prior Convictions	--	--	--	0.19	0.01	0.01
3 – Severity of Prior Convictions	0.09	0.02	0.03	--	0.05	0.15
4 – Escape History	--	0.02	0.01	0.07	--	--
5 – Past Institutional Behavior	0.36	0.20	0.27	0.45	0.22	0.21
6 - Age at First Conviction	0.03	0.18	0.31	0.41	0.22	0.31
7 – Current Age	0.47	0.44	0.64	0.12	0.19	0.27

8 – Stability Factors	0.27	0.25	0.35	0.47	0.30	0.37
9 – Projected Length of Incarceration	--	--	--	--	--	--
Model AUC	0.69	0.67	0.73	0.70	0.63	0.68

Two important findings are revealed in these optimized Hardyman models. First, most (but not all) of the Hardyman items were found to assist in the prediction of the three infractions outcomes. This is true for both males and females. Second, the optimized weights generated by the survival models substantially increase prediction model performance. Specifically, two male models (Non-Serious & Serious) were identified to possess "strong" prediction strength. The remaining models meet the "moderate" prediction threshold. Using this new weighting system alone, the Hardyman models would be drastically improved.

New Initial Classification Models

To further improve predictive performance, the item pool was expanded (see Appendix I) and new models were constructed. Six additional survival regression models were computed. Model findings are presented in Table 5. Again, AUC statistics indicated improved performance with the added items and weighting schematic. Each model improved upon the Hardyman models previously presented. Notably, four of the six models were identified as possessing a strong effect size (AUC=0.71-0.74), and the reaming two were of moderate strength (AUC=0.68).

It is also important to note that many of the original Hardyman items are consistent predictors of all six models. The removal of several Hardyman items may be the result of more predictive items included that are more precise, yet similar in content. With that said, several of the new items introduced were not selected and did not improve model prediction; these items have no coefficient values in the table below.

Table 5. Survival Regression Coefficient Initial Classification Estimates

	Male			**Female**		
Item	**Violent**	**Serious**	**Non-Serious**	**Violent**	**Serious**	**Non-Serious**
Security Threat Group Verified	0.54	0.40	0.45	0.39	0.39	0.69
Sentenced for Felony						0.03
Current offense						
Violent				0.08		
Property	0.11	0.13	0.30		0.19	0.05
Drug						
Sex						
Other						
Age At Admission - R	0.49	0.45	0.70	0.45	0.44	0.35
Age of First Conviction - R						
Highest Grade Completed - R						
Full Time Employment/Child Care - R						
Earliest Release Date				0.05		
Detainer Issue	0.35	0.08			0.10	
Prior Incarcerations	0.01	0.01		0.20	0.02	
Seriousness of Current offense						
Prior Violent Offense	0.01	0.01		0.03		
Prior Property Offense	0.03	0.03	0.04	0.03	0.02	0.02
Prior Drug Offense						
Prior Sex Offense	0.03					
Prior Weapons Offense	0.08		0.07	0.17	0.04	0.12
Prior Escape Offense				0.58		0.19

Prior Domestic Violence Offense			0.02	0.05		0.14
Prior Assault Offense	0.09			0.22		
Prior Violent-Property Offense		0.02		0.12		
Prior Any Offense						
Prior Incarceration Any Infractions		0.01	0.01		0.01	0.01
Prior Incarceration Violent Infractions	0.15		0.14	0.04		0.44
Prior Incarceration Serious Infractions		0.01			0.01	
Discretionary Overrides						
D1 – Multiple (4+) FTAs						
D2 – Pending Investigation						
D3 – Program Participation	0.03	0.09			0.09	0.09
D4 – NDCS Need	0.35	0.32	0.12	0.54	0.32	0.06
D5 – Medical Conditions						
D6 – Protective Custody	0.84	0.20	0.42	1.51	0.20	0.76
D7 – Mental Health Concern						
D8 – Non-Compliance with Program Rules						
D9 – Central Monitoring						
D10 – Security Threat Group						
D11 – Active Detainer/Pending Charges						
D12 – Escape Threat						
D13 - Other						
Mandatory Overrides						
M1 – Death Penalty Case						
M2 – Detainer with Fines Only						
M3 – ICE Detainer	0.39					
M4 – Low Severity Detainer			0.17			0.12
M5 – Moderate/High Severity Detainer	0.01	0.03	0.05	0.55	0.04	0.10
M6 – Tentative Release Date 8+ Years						
M7 – Tentative Release Date 5+ Years						
M8 – Tentative Release Date 3+ Years						
M9 – *Administrative Confinement*						
Hardyman Items						
1 – Severity of Current Offense	0.07	0.04	0.09	0.15	0.04	0.10
2 – Number of Prior Convictions						
3 – Severity of Prior Convictions	0.07	0.02	0.06	0.03	0.02	0.15
4 – Escape History				0.04	0.01	
5 – Past Institutional Behavior						
6 - Age at First Conviction	0.04	0.12	0.17	0.19	0.14	0.22
7 – Current Age						
8 – Stability Factors	0.23	0.21	0.29	0.41	0.22	0.32
9 – Projected Length of Incarceration						
Model AUC	0.71	0.68	0.74	0.73	0.68	0.71

R – Indicates an item was reverse coded

The items and weights identified here represent the preliminary models. Each item weight represents a score to be used on a linear scale, providing the point value indicated for each increasing response. These item scores are then summed to create an overall risk score for each infraction type. These scores are then divided to create risk categories that combine all three infraction model types. Classification scoring forms are provided in Appendix II.

Initial Classification Cut points

Next, risk categories were established by setting cut points for each model. As described, high risk cut points were set for each of the three models, where the higher severity category was given priority if an offender was identified to be of high risk in more than one model. Those not exceeding the high risk cut point in any model were identified as low risk. The initial cut point hazard ratios, category percentages and the estimated hazard rate of each event are provided in Table 6.

Table 6. Initial Classification Category Proportions and Infraction Rates

Category	Cut Point HR	Category %	Violent HR	Serious HR	Non-Serious HR
Female					
1. High Violent	4	12	7	3	5
2. High Serious	3	8	5	2	3
3. Moderate	3	14	2	2	2
4. Low	--	66	--	--	--
Male					
1. High Violent	4	18	5	4	6
2. High Serious	3	5	2	3	4
3. Moderate	3	12	2	2	3
4. Low	--	65	--	--	--

To establish cut point locations, infraction hazard ratios were used to identify the propensity of an offender in the high risk group failing, as compared to the remainder of the sample (or those not in the high risk group). The high risk hazard ratio cut point was set to roughly four for violent infractions and three for serious and non-serious infractions[5]. The particular hazard ratios values were selected as they approximated the current custody levels percentages of the sample and they provided an intrinsic value. For example, a hazard ratio of four indicates that those offenders classified as High Violent possess four times the odds of committing a violent infraction in the six months following admission when compared to all other offenders in the sample.

In the next column, the category percentages are displayed. These percentages represent the proportions of the sample that are identified to be in each category, following the hierarchical classification system. It should be noted that roughly 20% of females were identified to be either High Violent or Serious, which is comparable to the roughly 26% of the female population that received an approved custody level designation of either Maximum or Medium (see Table 1). Similarly, the combined 80% identified as Moderate or Low risk is comparable to the 74% of the female population at the minimum or community custody level. This same finding is observed for males, where the combined 23% of High Violent and Serious offenders are similar to the 22% approved for maximum or medium custody.

In the final three columns, hazard ratios are provided for each infraction category as it pertains to each infraction type. What one can observe is that, despite the tailoring of categories to a specific infraction type, those offenders in a higher risk category possess a greater propensity for failure than those in a lower risk category. For example, compared to low risk offenders, High Violent female offenders possess five, three and seven times the propensity of commit non-serious, serious and violent infractions following their initial classification; where male High Violent offenders possess sex, four and five times the propensity for each infraction type, respectively.

Collectively these findings indicate that category cut points were placed to provide substantial discrimination between infraction group types. Furthermore, estimates suggest that population percentages

[5] It should be noted that hazard rates were rounded to the nearest whole number to provide a more easily interpretable table but may be a 1-4 tenths different than displayed.

will be similar to those of the current approved custody designations. The improved accuracy of the models, indicated by large model AUC values (Table 5), indicate the custody designation will place offenders in the appropriate category more accurately. Finally, the hierarchical classification system has the ability to identify those at highest risk of a particular infraction and all infraction generally.

Reclassification

Next, using the reclassification sample reclassification infraction models were created, replicating the methods for the development of the initial classification. While the three outcomes remained the same, the pool of potential items to be included in the new models did expand (see Table 2). In addition, several of the Hardyman items differ slightly at reclassification, taking into consideration behavior that has been observed since admission, including: involvement with drugs or alcohol, disciplinary infraction frequency and severity, and performance in work/programming. Again, we examined infraction prediction validity using the Hardyman summary scores, models selecting only the Hardyman items, and created new models with the expanded prediction item pool.

Hardyman Scores and Models

Using the Hardyman reclassification items, summary scores were calculated for each offender. Again, these scores represent those calculated by NDCS assessment staff. The summary scores were then entered into each one of the survival regression models. AUC statistics for each model are presented in Table 7. As identified in the initial classification, only one model (Male Violent) identified to possess a moderate effect size (AUC=0.65). Furthermore, only one additional model (Female Violent) identified a small effect size (AUC = 0.62), while all the remaining models identified negligible effect sizes. This is an alarming finding, where four of the six models possessed equal or slightly improved accuracy over random chance. With this said, the Handyman models were not operationalized to predict these three specific outcomes, however, one would anticipate model effects that were at least "small" by industry standards. A positive takeaway from these findings would suggest that new models developed possess a great potential to improve predictive performance.

Table 7. Hardyman Score and Survival Reclassification Infraction AUC

Model	Violent	Serious	Non-Serious
Female	0.62	0.50	0.50
Male	0.65	0.53	0.51

As described in the initial classification discussion, the Hardyman reclassification models were not developed with currently available methodological techniques and were not construct models to predict the specific outcomes identified here. The individual items were then added to their own prediction models to select and optimally weight each prediction item for each outcome. The customized selection procedure was again used to select and weight items using survival modeling. The resulting model items, weights, and AUC statistics are provided in Table 8.

Table 8. Hardyman Items and Survival Reclassification Infraction AUC

	Male			Female		
Item	**Violent**	**Serious**	**Non-Serious**	**Violent**	**Serious**	**Non-Serious**
1 – Involvement with Drugs or Alcohol		0.10				0.02
2 – Escape History	0.01		0.04			

3 – Frequency of Disciplinary Infractions	0.08			0.10	0.05	
4 – Severity of Disciplinary Infractions	0.14	0.05	0.02	0.15		0.05
5 – Conviction History	0.08			0.36		
6 – Current Age	0.29	0.25	0.32	0.45	010	0.04
7 – Performance in Work/Programming	0.23			0.06	0.08	0.20
Model AUC	0.69	0.56	0.57	0.72	0.53	0.57

One can observe two important findings regarding the optimized Hardyman reclassification models. First, many of the Hardyman items were found to assist in the prediction of the three infractions outcomes. This was again true for both males and females. Second, the optimized weights generated by the survival models substantially increases infraction prediction model performance. Specifically, the Female Violent model was identified to possess a strong effect size (AUC = 0.72) and the Male Violent model was found to possess a moderate effect size (AUC = 0.69). With regard to the remaining models, three were identified to possess a small prediction strength (AUC=0.57), while only one model was found to be of negligible strength (AUC = 0.51). Again, using this new weighting system alone, the Hardyman models would be substantially improved.

New Reclassification Models

Using the more advanced item selection and weighting methods, new models were constructed predicting infractions following a reclassification. To further improve predictive performance, the item pool was again expanded (see Table 2). Six additional survival regression models were computed. Model findings are presented in Table 9. Again, AUC statistics indicated improved performance with the added items and weighting schematic. Based on industry standards, two of the six models possessed moderate strength (AUC=0.64 & 0.66) while four identified as having strong effect sizes. It should be again noted that the items and weights identified here represent the preliminary models (AUC = 0.76-0.77). Each item weight represents a score to be used on a linear scale, providing the point value indicated for each increasing response[6].

Table 9. Survival Regression Coefficient Reclassification Estimates

	Male			**Female**		
Predictors	**Violent**	**Serious**	**Non-Serious**	**Violent**	**Serious**	**Non-Serious**
Security Threat Group - Verified	0.58	0.01				
Sentenced for Felony	0.01	0.10		0.06	0.38	0.16
Current offense (not mutually exclusive)						
Violent						
Property	0.23	0.29		0.48		
Drug		0.21				
Sex				0.45		
Other						
Age At Reclassification	0.58	0.30	0.25	0.24	0.19	0.50

[6] It should be noted that tests of independence indicated that the male sample violated the regression assumption but the female sample did not; therefore, the analysis of the female sample allowed for multiple cycles (multiple reclassification assessments from the same offender) while the male sample removed multiple cycles, randomly.

Age of First Conviction						
Highest Grade Completed			0.02	0.48	0.08	0.30
Full Time Employment/Child Care Prior						
Earliest Release Date				0.11		
Detainer Issue	0.01				0.20	
Prior NDCS Incarcerations	0.10				0.21	
Seriousness of Current offense						
Prior Violent Offense				0.09	0.04	
Prior Property Offense		0.04	0.04	0.04		
Prior Drug Offense		0.02			0.01	
Prior Sex Offense	0.08		0.03		0.04	
Prior Weapons Offense	0.04	0.10	0.05	0.61	0.03	0.09
Prior Escape Offense		0.09			0.53	0.36
Prior Domestic Violence Offense		0.18	0.22			0.11
Prior Assault Offense	0.15	0.03	0.03	0.01	0.06	0.02
Prior Violent-Property Offense	0.17					
Prior Any Offense						
Prior Incarceration Any Infractions						
Prior Incarceration Violent Infractions						
Prior Incarceration Serious Infractions						
Any Infractions During Current Incarceration	0.01			0.44		
Violent Infractions During Current Incarceration	0.08	0.12			0.08	
Serious Infractions During Current Incarceration						
Violent/ Serious Infractions During Current Incarceration	0.23			0.19		
Any Infractions 6 Months Prior to Reclassification	0.21	0.22	0.24	0.01	0.09	0.76
Violent Infractions 6 Months Prior to Reclassification	0.07					
Serious Infractions 6 Months Prior to Reclassification						
All Prior Any Infractions			0.01		0.02	
All Prior Violent Infractions	0.18					
All Prior Serious Infractions		0.05			0.31	0.14
Incarceration Duration to Reclass						
Discretionary Overrides						
D1 – Multiple (4+) FTAs						
D2 – Pending Investigation						
D3 – Program Participation		0.01		0.34	0.12	0.21
D4 – NDCS Need	0.56	0.17			0.66	
D5 – Medical Conditions						
D6 – Protective Custody	0.39	0.24	0.11			
D7 – Mental Health Concern						

D8 – Non-Compliance with Program Rules	0.89	0.29	0.41	0.83	0.90	0.66
D9 – Central Monitoring/Separation Issues						
D10 – Security Threat Group						
D11 – Active Detainer/Pending Charges						
D12 – Escape Threat						
D13 – Other						
Mandatory Overrides						
M1 – Death Penalty Case						
M2 – Detainer with Fines Only						
M3 – ICE Detainer	0.82					
M4 – Low Severity Detainer		0.03	0.18	0.72		
M5 – Moderate/High Severity Detainer	0.44		0.06	0.85	0.04	
M6 – Tentative Release Date 8+ Years	0.13					
M7 – Tentative Release Date 5+ Years	0.07			0.75	0.22	
M8 – Tentative Release Date 3+ Years						
M9 – *Administrative Confinement*		0.02				
Hardyman Items – R						
1 – Involvement with Drugs or Alcohol	0.11	0.10	0.04			0.09
2 – Escape history	0.04					
3 – Frequency of Disciplinary Infractions						
4 – Severity of Disciplinary Infractions						
5 – Conviction History	0.12			0.17		
6 – Current Age						
7 – Performance in Work/Programming	0.06	0.03	0.01	0.09	0.07	0.13
Reclass Facility Location						
Facility Location Following Reclass						
Transferred to a New Facility Following Reclass						
Model AUC	0.77	0.66	0.76	0.76	0.64	0.76

Unlike the initial classification models, the AUC prediction is not identified to be as consistent or universally strong. This may be due to the removal of several static predictors (i.e., criminal history) from the item pool, which demonstrated improved model accuracy through the inclusion of a larger prediction item pool. However, the reduced model AUC between initial and reclassification assessments may also be due to a lack of dynamic items that reflect improvement offenders' current attitudes and characteristics impacting infraction behavior. The potential improvement gained via greater dynamic items will be discussed in greater details in the final section.

Reclassification Cut points

To establish reclassification cut point locations, infraction hazard ratios were again used to identify the propensity of an offender in the high risk group failing, as compared to the remainder of the sample (or those not in the high risk group). As the reclassification assessment has the potential to substantially alter the

proportions of offenders assigned to each custody level, the particular hazard ratio values were selected to approximate the current custody levels percentages of the sample. Presented cut points are considered preliminary and may be altered based on SME input. Cut point HRs, Category percentages, and infractions specific HRs are presented in Table 10.

Table 10. Reclassification Category Proportions and Infraction Rates

Category	Cut Point HR	Category %	Non-Serious HR	Serious HR	Violent HR
Female					
1. High Violent	7	14	9	3	13
2. High Serious	3	28	3	2	3
3. Moderate	3	11	3	2	2
4. Low	--	47	--	--	--
Male					
1. High Violent	5	16	9	3	7
2. High Serious	3	23	7	2	3
3. Moderate	3	21	3	2	2
4. Low	--	40	--	--	--

In the second column, the category percentages are displayed. These percentages represent the proportions of the sample that are identified to be in each category, following the hierarchical classification system. It should be noted that roughly 42% of females were identified to be either High Violent or Serious, which is comparable to the roughly 41% of the female population that received an approved custody level designation of either Maximum or Medium (see Table 2). This same finding is observed for males, where the combined 39% of High Violent and Serious offenders are similar to the 37% approved for maximum or medium custody.

In the final three columns, hazard ratios are provided for each infraction category as it pertains to each infraction type. Again, one can observe is that those offenders in a higher risk category possess a greater propensity for failure than those in a lower risk category. For example, compared to low risk offenders, High Violent female offenders possess nine, three and thirteen times the propensity of commit non-serious, serious and violent infractions following their initial classification; where male High Violent offenders possess nine, three and seven times the propensity for each infraction type, respectively. Similar to the initial classification, these findings indicate that category cut points were placed to provide substantial discrimination between infraction group types, the population percentages are similar to current custody designations, and the hierarchical classification system has the ability to identify those at highest risk of a particular infraction and all infraction generally. It should be noted that the HR for violent offenses were set higher than that of the initial classification models as these infraction types occur with less frequency in the reclassification samples.

Classification and Risk Category Interactions

The new risk categories were designed to indicate the infraction type that is most likely for a given offender. This is a departure from the Hardyman tools, in that the new models do not specify where the offender should be placed. This, in turn, allows staff to make recommendations based on security needs surrounding infractions behavior but also allows flexibility when bed space needs may require staff promote or demote based on an offender's risk relative to other offenders. To assist with this decision, a final set of analyses was computed to examine the offender's propensity for failure in a given custody level, based on their scored risk category.

Hazard ratios (HRs) were computed, broken down by custody level. Based on responses collected during the process evaluation indicating that maximum and medium security generally provide the same

security restrictions, these custody designations were merged into a single category for the purposes of the analyses. The HRs for offenders overall were first computed and then a breakdown of HRs by risk level is provided. To simplify the presentation of findings, the Low and Moderate Risk groups and the High Serious and Violent risk categories were merged. For all analyses, the community custody designation is used as a reference group, comparing offenders with this designation to minimum and max/medium.

Findings for the female initial classification analyses are provided in Table 11. Overall, compared to females assigned to community custody, those assigned to max/medium custody possessed three times the hazard of committing a non-serious infraction, while those assigned to minimum possessed twice the odds. When examining serious infractions, females assigned to max/medium possessed twice the hazard of failure, while those assigned to minimum possessed even odds of failure when compared to community placements. With regard to violent infractions, compared to community placements, those assigned to max/medium possessed seven times the hazard, while those assigned to minimum possessed twice the odds.

Table 11. Initial Classification Infraction Hazard Ratio – Female

Category	Non-Serious	Serious	Violent
Overall			
Max/Medium	3	2	7
Minimum	2	1	2
Community (reference)	--	--	--
Low/Moderate			
Max/Medium	2	2	5
Minimum	1	1	2
Community (reference)	--	--	--
High Serious/Violent			
Max/Medium	2	2	3
Minimum	2	1	2
Community (reference)	--	--	--

When examining females scoring as low/moderate risk, compared to those placed in community facilities, those offenders assigned to max/medium possess twice the hazard for committing a non-serious or serious infraction, while those assigned to minimum possess roughly even odds. With regard to violent offenses, those assigned to max/medium possess five times that hazard and those in minimum have twice the hazard of those placed in the community.

Examining females scoring as high risk, compared to those placed in community facilities, those offenders assigned to max/medium or minimum possess twice the hazard for committing a non-serious or serious infraction, while those in minimum possess roughly even odds as those in the community facilities. With regard to violent offenses, those assigned to max/medium possess three times that hazard and those in minimum have twice the hazard of those placed in the community.

Findings for the male initial classification analyses are provided in Table 12. Overall, compared to males assigned to community custody, those assigned to max/medium custody possessed four times the hazard of committing a non-serious infraction, while those assigned to minimum possessed twice the odds. When examining serious infractions, males assigned to max/medium possessed three times the hazard of failure, while those assigned to minimum possessed even odds of failure when compared to community placements. With regard to violent infractions, compared to community placements, males assigned to max/medium possessed eight times the hazard, while those assigned to minimum possessed three times the odds.

Table 12. Initial Classification Infraction Hazard Ratio – Male

Category	Non-Serious	Serious	Violent
Overall			
Max/Medium	4	3	8
Minimum	2	1	3
Community (reference)	--	--	--
Low/Moderate			
Max/Medium	4	2	8
Minimum	1	1	3
Community (reference)	--	--	--
High Serious/Violent			
Max/Medium	3	2	5
Minimum	2	2	3
Community (reference)	--	--	--

When examining males scoring as low/moderate risk, compared to those placed in community facilities, offenders assigned to max/medium possess four and two times the hazard for committing a non-serious or serious infraction (respectively), while those assigned to minimum possess roughly even odds. With regard to violent offenses, males assigned to max/medium possess five times that hazard and those in minimum have three times the hazard of those placed in the community.

Examining males scoring as high risk, compared to those placed in community facilities, those offenders assigned to max/medium or minimum possess four times the hazard for committing a non-serious and twice the odds of committing a serious infraction, while minimum custody placements possess roughly even odds for both infraction types. With regard to violent offenses, those assigned to max/medium possess five times that hazard and those in minimum have three times the hazard of those placed in the community.

Findings for the female reclassification analyses are provided in Table 13. Overall, compared to females assigned to community custody, those assigned to max/medium custody possessed seven times the hazard of committing a non-serious infraction, while those assigned to minimum possessed three times the odds. When examining serious infractions, females assigned to max/medium possessed twice the hazard of failure, while those assigned to minimum possessed roughly even odds of failure when compared to community placements. With regard to violent infractions, compared to community placements, those assigned to max/medium possessed eight times the hazard, while those assigned to minimum possessed four times the odds.

Table 13. Reclassification Infraction Hazard Ratio – Female

Category	Non-Serious	Serious	Violent
Overall			
Max/Medium	7	2	8
Minimum	3	1	4
Community (reference)	--	--	--
Low/Moderate			
Max/Medium	3	2	9
Minimum	3	1	3
Community (reference)	--	--	--
High Serious/Violent			
Max/Medium	6	1	5
Minimum	3	1	2
Community (reference)	--	--	--

When examining females scoring as low/moderate risk, compared to those placed in community facilities, those offenders assigned to max/medium or minimum possess a hazard that is three times that of females in community placements. For serious infraction, while those assigned to max/medium possess twice the hazard, while minimum custody offenders possess roughly even odds. With regard to violent offenses, those assigned to max/medium possess nine times that hazard and those in minimum possess a hazard that is three times that of those placed in the community.

Examining females scoring as high risk, compared to those placed in community facilities, those offenders assigned to max/medium or minimum possess three times the hazard for committing a non-serious infraction. For serious infractions, those assigned to max/medium possess twice the hazard, while those in minimum possess roughly even odd as those in the community facilities. With regard to violent offenses, those assigned to max/medium possess five times that hazard and those in minimum have twice the hazard of those placed in the community.

Findings for the male reclassification analyses are provided in Table 14. Overall, compared to males assigned to community custody, those assigned to max/medium custody possessed twice the hazard of committing a non-serious infraction, while those assigned to minimum possessed roughly even odds. When examining serious infractions, males assigned to max/medium or minimum custody possessed roughly even odds of failure when compared to community placements. With regard to violent infractions, compared to community placements, those assigned to max/medium possessed four times the hazard, while those assigned to minimum possessed twice the odds.

Table 14. Reclassification Infraction Hazard Ratio – Male

Category	Non-Serious	Serious	Violent
Overall			
Max/Medium	2	1	4
Minimum	1	1	2
Community (reference)	--	--	--
Low/Moderate			
Max/Medium	2	1	3
Minimum	1	1	2
Community (reference)	--	--	--
High Serious/Violent			
Max/Medium	1	1	3
Minimum	1	1	1
Community (reference)	--	--	--

When examining males scoring as low/moderate risk, compared to those placed in community facilities, those offenders assigned to max/medium possess a hazard that is twice that of offenders in community placements, while minimum custody males possess roughly even odds. For serious infraction, while those assigned to max/medium, or medium possess roughly even odds when compared to community custody offenders. With regard to violent offenses, those assigned to max/medium possess three times that hazard and those in minimum possess a hazard that is twice that of those placed in the community.

Examining males scoring as high risk, compared to community placements, those offenders assigned to max/medium or minimum possess roughly even odds of committing a non-serious or serious infraction. With regard to violent offenses, those assigned to max/medium possess three times the hazard and those in minimum or community facilities.

Custody-level Trends

Taking the analyses presented in the four tables, one can observe that, collectively, infractions of any type are at their lowest in community facilities. However, this is not to suggest that all offenders should be housed in community facilities to prevent infractions. Currently the classification staff members (in conjunction with the Hardyman tools) are placing some of the highest risk offenders in higher custody designations and those offenders are most likely to commit misconduct. Also, the facility and the rate of infractions have an impact, where greater restrictions on movement and a higher staff-to-inmate ratio combine to result in the occurrence and greater observation of infractions.

With this said, an interesting finding is observed when examining the HRs of low/moderate risk offenders placed in max/medium custody. For these offenders, the odds of committing a violent infraction is equal to and often greater than that of the overall sample and that of high risk offenders. A central concept of Andrews and Bonta's Risk, Need, and Responsivity (RNR) model is that lower risk offenders can be observed to commit a greater rate of negative behavior when grouped with high risk individuals. While their model was primarily concerned with recidivist tendencies, this same pattern is observed when offenders identified to be of low risk to commit and infraction are placed in a facility that houses a greater proportion of high risk offenders and is highly restrictive.

Section Summary

Overall this section described the current validity of the Hardyman tools, the development of new models and the comparisons of each model's predictive validity. Overall models were constructed as planned and findings were found to indicate successful improvements. The currently utilized Hardyman classification scores were found to provide weaker levels of predictive accuracy when compared to the new models. These findings were anticipated as more advance statistical techniques were used to select items and weight response scores. The additional number and specificity of predictor items also assisted in improved prediction capacity of the models created.

Cut points for each model were then established, identifying those offenders that are higher risk to commit a given infraction type. Following classification, offenders were then classified into one of four infraction risk categories. Hazard ratios for each category were then presented, identifying an offender's propensity to commit an infraction based on their defined risk category.

Finally, tables of infraction hazard ratios describing and comparing placement of the various infraction risk levels within given custody designations. Generally, the findings provided in these tables are meant to be informative. The classification system created from the new models does not provide a recommended custody placement and instead outlines the level of risk and the infraction type an offender is most likely to commit. Intuitively classification staff members will most likely place high risk offenders in max/medium, while low and moderate risk offenders will be assigned to minimum and community facilities. However, several NDCS policies and procedures will prevent custody assignment to be based solely on an offender's risk category. Some examples of these policies and procedures may include: an offender's reentry/releasing facility, mandatory overrides, and bed space needs/restrictions. When classification staff are faced with these decisions, these tables may serve as a reference, identifying the relative risk of placing an offender in a given risk category in a particular custody level. In the next section SME contributions and modifications to the final models are described.

IV – Subject Matter Expert (SME) modifications and considerations

On July 26th preliminary model findings were presented. First, a brief presentation was provided to Director Frakes and his executive team members for their Tuesday staff meeting. To outcomes and the findings of the preliminary tool were generally viewed positively. The discuss mostly centered around the use of the tool going forward, potential changes in policy, and the need to create implementation and training materials. Timelines were also conveyed, with a desired need to beginning training and implementation prior to September. A four hour SME meeting was then facilitated, discussing the development of the tool and the infraction model findings.

Operational definitions

Feedback around the tool and the newly included items were generally positive. However, SMEs raised several concerns. First, there were some minor concerns regarding the operational definitions of the infraction predictors and outcomes. SMEs noted that several infraction offense classifications have changed over the years. Furthermore, several infractions offenses have been eliminated and added over the last 20 years. While these modifications possess a minor impact when considering the relatively consistent scoring of most infractions over time, further investigation is warranted.

Additional labor

Another concern was the new tools' impact on labor. Several items have been added to both the initial classification and reclassification models. Data collection will require a greater duration of assessment staff completing the scoring. It was suggested that some of this labor could be mitigated through the use of automation, using software to populate redundant fields. Specifically, the classification tools use many items that are similar in content to the STRONG-R risk and needs assessments tool currently in use. Staff suggested that the inclusion of the classification tools in the STRONG-R software would speed up assessment labor at initial classification, eliminating redundancies in data collection. With regard to reclassification, several predictor items are static, criminal history measures, which will not likely change during an offenders' incarceration. Therefore, a method to take collected information forward from initial to reclassification would also reduce labor demands.

Policy impact

A key concern was the impact that the new tools will have on policy. Efforts were made with the current infractions models to include mandatory and discretionary overrides as predictors. The intent was to eliminate, or greatly reduce, the use of modeled overrides. SMEs were concerned about the possibility of removing specific overrides entirely. That said, SMEs were also concerned that the retention of all overrides would result in a similar rate identified in prior version of the classification instruments (i.e., 40%).

In addition, staff members were also concerned with the Department's policy/goal of returning offenders from a community facility. While releasing offenders form a community facility was intended to increase opportunities for those offenders with limited social supports or those needing reentry services (i.e., employment, residence, family) the policy seemed to be applied more universally than intended. Given the lack of bed space in community facilities and additional security and misconduct concerns, the policy direct goal did not seem entirely achievable. Specifically, classification SMEs identified a bottleneck of offenders in maximum and medium security. Where open beds are often available at minimum custody level designations. Programming is a big concern with many offenders, as clinically recommended interventions (i.e., substance abuse treatment, violence reduction, and sex offender treatment) have long waiting lists and are not available in all facilities. Offenders are often not considered for a transfer to a lower custody designation until a parole date is set (roughly 12 months in advance). It was conveyed that parole is reluctantly granted unless the

offender has completed needed clinically recommended programming and said programming is not offered in community facilities. Furthermore, protective custody is not available at all facilities, preventing the graduation of custody designations for special populations (i.e. sex offenders). Because of these intertwining issues, the research division was asked to investigate and found that only 4 in 10 offenders are discharged from a community facility[7]. Due to current limitations of data collected, it was not determined how may offenders discharging from community facilities were of "high need" or lacked reentry support systems.

A related complication was expressed with regard to offenders' TRD. While the bottleneck of offenders was apparent in maximum and medium security facilities, classification staff members were reluctant to transfer offenders with longer sentences (8 or more years) to a minimum security facility. SMEs discussed how some longer term offenders would be appropriate for minimum custody designation, based on their relative low risk for infractions. However, classification staff members were also hesitant to make such placements as it may shift the bottleneck to minimum security, potentially delaying the promotion of shorter TRD offenders due to a longer term offender holding a needed bed.

Overall there were several conflicting policies and departmental needs related to the current and new tool implementation and use. The issues regarding offenders' TRD, an offender's risk to infract, and the stated policy goal of releasing offenders from community facilities has created a conflict that requires a shift in policy. To be discussed more in the recommendations section, the NDCS will need to revisit the need for current overrides and potential remove some or update current polies to adjust for previously state needs regarding over-classification, bed space and bottleneck issues, and graduated promotion in custody levels.

Retention of key overrides

SMEs also discussed the need to retain several overrides. Specifically, those with a serious mental health issues may not be eligible to transfer away from special or segregated custody. Furthermore, mental health staff are needed "nearby" incase such an offender destabilizes. In particular, it was expressed that overrides be used for placements of severely mentally ill offenders in WEC, as there are no mental health staff members in the relative proximity. A related override is needed for seriously mentally ill and sex offenders. Currently, offenders are required to obtain a positive psychiatric evaluation in order to be promoted. SMEs indicated that a need to retain this policy and used it within the override options going forward. Offenders in need of protective custody should also maintain an override option. This override will continue to retain the safety and well-being of offenders and staff.

With regard to community custody, an override relating to detainers should be retained. While detainers related to misdemeanors, fines, and other costs could (and many thought should) be removed from the list of override options, serious detainers should be maintained as an override option to prevent walk-ways from community facilities. It was suggested that Class I offenses be considered a serious detainer criterion but more discussion is likely needed to approve this override policy change.

Lastly, diagnostic SMEs indicated that scoring of the Hardyman tool regarding the "escapes" predictor crated an issue for parole violators. This item was retained in the new models, as it was found to be a good predictor of future infraction behavior. The issue relates to the potential over-classification of offenders returning on a parole violation. Many of these offenders were recently released from a minimum or community custody facility. However, the score of the escape measure upon return to custody typically places an offender in maximum or medium. Without investigating this issue further prior to and during implementation, it is difficult to know if the scoring issues around parole violators will continue to be an

[7] This estimate was based on 1,572 offenders released in the 2016 Fiscal Year.

issue. However, it may be important to consider an additional override as it pertains to these cases, limiting the impact of parole violations as it pertains to escapes.

Pilot needed

Discussions with SMEs also revealed concerns with a quick implementation of the new tool. The general concern being that the tool may identify offenders to be over or under-classified or that particular items might not function as anticipated. A typical process is a testing and pilot phase of any new tool. In our final section we will discuss this recommendation and provide more details outlining our suggested approach.

Scoring Examples

SMEs were then presented with the preliminary scoring guide for the tools. They were tasked with examining the tool for content and terminology. In addition, SMEs were asked to calculate scores on the new tool using a few current cases as examples. It was requested that a detailed description of their perception of the process be provided and any potential issues they could foresee following implementation. A total of eight offender scoring sheets were completed and returned. Also, Hardyman category scores were also provided as a comparison

Examining the general themes of SME's reports indicated that the scoring guide needed to be improved. The initial draft did not provide sufficient detail and will need greater specificity when developed to training materials for staff. Example cases that discuss how to count prior offenses and infractions were identified as potentially beneficial. Following implementation and/or pilot testing, the definition and scoring of each item may need to be adjusted to improve accuracy and tool clarity.

A second concern was with regard to the duration of time the new tools took to complete. SMEs indicated that the extensive document review of offenders' criminal history and corrections events substantially extended the duration of time needed to complete each assessment. This was of particular concern for reassessments, where the Hardyman tools used only seven items require little time to complete. Furthermore, some items, such as those counting the number of prior infractions, take a substantial amount of additional time but some may not weigh that heavily in the final score. While the extended duration may initially seem defeating, other SMEs noted that training and practice will likely improve efficiency over time. Other staff noted that the ability to use data collected via the recidivism risk assessment tool (the STRONG-R) would also improve the speed at which the tool could be completed.

A final concern was with regard to the categories in which the offenders scored. On the lower ends of the spectrum, those scoring "low" tended to also score as "minimum" or "community" on the Hardyman scales. However, those scoring as "maximum" on the Hardyman scales had a tendency to score lower on the new tool. This is a concerning finding and will need to be flowed up with great test cases or through a pilot study. If the issue continues to present itself in additional test cases, cut points may need to be adjusted to assure that the appropriate offenders are identified for higher custody level designations.

Section Summary

This section outlined findings received from requested SME feedback on the new tools created. This is considered one of many opportunities for staff to assist in the development and further refinement of the classification tools and processes going forward. The current feedback provided indicated that great care should be taken with regard to item level definitions and training materials crafted. SEMs are generally concerned that the new tool, although, identified to improve accuracy, may create additional costs of assessment labor. Many staff members indicated that labor efficiencies would be made if it was feasible for the NDCS to merge efforts of software creation and/or information sharing around the STRONG-R tool with classification assessments. Finally, policies and practices will need to be examined and potentially

modified to merge with the new classification tool design and to eliminate those policies and goals that are at odds with current NDCS needs.

VI – Conclusions & Recommendations

During the course of this project, several phases were used to collect and analyze data to improve performance, process, and usage of the NDCS classification system. In Phase II a document review, interviews, and focused groups outlined several issues with the usage of the current tools and issues related to current classification processes were outlined. While several substantial strides were made to address identified needs, it was not feasible to address the totality of issues outlined by NDCS SMEs. The final phase of the project is used to outline issues that were resolved or mitigated with the current efforts and those that remain.

Classification Tool Improvements

As described in Part II of the report, several methodological limitations were observed with the Hardyman classification tools. Briefly, identified issues included: a small development sample, a lack of prediction item specificity, unweighted scoring system, a less-than-optimal feature selection procedure, inconsistent item usages between classification and reclassification tools, a lack of infraction outcome specificity, less-than-efficient validation techniques, inability to account for offenders' infraction behavior within a given facility/custody designation. The current modeling efforts attempted to adjust for these issues and were successful in making several improvements.

First, substantial improvements were made to the sample sizes of both the initial and reclassification samples for both males and females. All samples sizes exceeded minimums indicated by prior research (Styerberg et al., 2001). In addition, extensive efforts were made to gather offender criminal history and institution data routinely collected by the NDCS, which greatly expanded the pool of potential items to be included in the classification tools. Efforts were also made to make all items available for both classification and reclassification samples. This was done to improve the consistency of item usage between the initial and reclassification tools' scoring.

Using SME input, three infraction outcomes were specified. Follow-up durations of each outcome were also operationalized to match the six-month prediction window between classification and reclassification assessments. To further improve the prediction accuracy, exposure times were assessed for each offender to identify their exact durations between assessments.

Advanced modeling techniques were also used to improve the predictive accuracy of created models. Feature selection techniques made use of a sophisticated survival regression method, which selected predictor items that improved the predictive accuracy. These techniques represent a customized solution, modified from those used in prior tool creations (Hamilton et al., 2016), and were optimized to fit the NDCS's classification processes. Specifically, these methods removed noise items that did not aid in the prediction of infractions and each item selected was given a weight customizing the tool's scoring to the specific characteristics and behaviors of the NDCS population. Finally, advanced validation techniques were used, which made use of the entire sample making item weights efficient and stable over time.

Item weights were then entered into a scoring sheet (see Appendix II). These scoring sheets represent the new tools to be used for classification decisions. Scores for each infraction type are summarized and model cut points determine an offender's risk category. These categories were developed to create distinctions between offenders with regard to their estimated propensity for infracting and also attempted to fit the current approved custody designations for the NDCS population. Unlike current classification tools, the risk categories provide users with a quantitative distinction of "high risk" and a qualitative understanding of the infraction risk type. Finally, information is provided as to the impact that a custody designation is likely to have on a given offender category. The cumulative impact of the outlined modifications has resulted in the creation of classification tools that are substantially improved when compared to the current instruments.

Addressing SME's suggestions

During interviews and focus groups, SME's provided many comments and suggestions regarding the current classification process. Some of these issues and suggestions could be feasibly addressed with the current project. For instance, several issues were raised regarding the scoring of the current tools, including: assessing the severity of misconduct, inconsistent scoring between initial and reclassification tools, the weighting of response items (resulting in over-classification), the use of predictor items that lacked face validity, and the long list of mandatory and discretionary overrides that could be included as scoring items within the current classification tools. Generally, staff indicated that the score computed by the current tools was not grounded in users' knowledge or was not designed to address NDCS classification policies and processes. Because overrides were so prevalent, the calculated score was described as a "suggestion".

Addressing scoring issues of the tool, we focus here on the strategic design of the outcomes, the follow-up durations, and the expansion of predictor items to be considered. The use of SME input helped frame the outcome definitions. Furthermore, the review of NDCS policies and processes led to the use of models that predicted infractions in six-month durations and adjusted for exposure times. With a comprehensive training and quality assurance (QA) plan, the scores and classification guidelines outlined by the tool should now be viewed as grounded in and defined by NDCS staff.

With regard to the expanded pool of prediction items, this effort was made to specifically address NDCS staff requests and was viewed as a likely source of untapped infraction prediction strength. In particular, prior criminal offenses and infraction types were specified to increase the accuracy and face validity of the tools. Both mandatory and discretionary override measures were added to the item pool in an effort to include these indicators as scoring items, and potentially removing their use as override factors going forward.

Finally, the new hierarchical categorization identifies the type of infraction that the offender is most likely at risk of committing. Therefore, the offender categories are now to be utilized as guidelines for classification, rather than a one-to-one recommendation as to where offenders should be placed. Informative tables are also provided to help users understand the risk associated with offender placements in particular custody designations. This new format will help inform classification recommendations, yet allow staff flexibility when bed space or other considerations are needed to guide decisions.

Recommendations

While substantial improvements were made with the development of the current classification tools, there were many staff needs and considerations that could not be addressed with the current project goals. The unaddressed issues raised by staff are not without merit but could not be easily addressed. Recommendations are provided for both short and long-term modifications for classification tools, processes, and future research.

Better training around the tool

There are two remaining phases of the project. Manuals and training materials will need to be developed to adjust the current classification tools and identify any updated policies and new procedures. Materials developed will guide training of new staff as well as refreshers for current staff. Goals for booster training and other quality assurance guidelines will also be developed. All materials will be created in conjunction with NDCS SMEs.

Following the development of training materials, an implementation plan should be established. The implementation plan is recommended to include a timeline for a graduated roll out of the new tools, a pilot study to assess system impact and address any modification needs to scoring and cut point placement, a training schedule, and outline for quality assurance checks and future validation analyses.

The worst result for a tool developer or an agency is to spend resources to create and implement an assessment, only to have the results ignored, or worse, viewed as a hindrance to day-to-day operations. The notion that staff members view the result of the Hardyman tool as a "suggestion" should be considered as a caution for future implementation efforts. The concept of implementing an assessment only to have its results diminished is not unique to NDCS classification. Recent research has suggested that in the absence of training and quality assurance QA procedures, assessment tools are not worth the labor and resources spent toward development (Viglione, Rudes, & Taxman, 2014).

The current project has outlined training efforts for the new implementation of the tools and adherence and continual updates of these processes will likely improve staff utilization and confidence. Training and implementation efforts will need to be diligent, informing staff of the rationales for the tool, the processes that went into creating the tool and its intended uses. Over time, the scoring and making of recommendations will become routine and hopefully more reliable with experience. However, QA procedures will need to assure that staff members do not become complacent, or drift, from the intent of the tool or the underlining definitions of the responses and outcomes.

Consistent boosters and refreshers of the tool will maintain staff connections with the results. Allowing staff to engage with the development of future versions and providing a medium to offer suggested improvement is also recommended. Classification decisions should not be made in a vacuum or be the result of a single item, tool, or person. Using well-trained staff to assess and provide recommendations based on evidence and experience will mitigate drift and improve decision-making practices.

Improving Prediction and Efficiencies

Despite expanding the number and specificity of prediction items used for the current assessment model development, those items used do not constitute the universe of potential predictive factors. In particular, many of the items used represent static, criminal, and institution items. An expanded list of dynamic, needs-based items would likely improve prediction, particularly for reclassification assessments. A tool developer will always want an agency to expand its pool of potential predictors but that expansion can come at a cost, creating increasingly longer assessments and greater labor demands. However, concerted efforts can be made to reduce redundancies and inefficiencies.

Specifically, during data collection for the process evaluation, staff provided documentation and discussed the detailed record collection for each offender. Diagnostic teams complete a battery of assessments for each offender upon admission. These psycho-social evaluations undoubtedly provide important information that helps inform the day-to-day diagnostic efforts but may also be of used for infraction prediction. Furthermore, the multitude of assessment tools used by the NDCS likely contains items with substantial overlap, collecting the same/similar content on multiple forms.

If a one-stop-shop for assessment data could be maintained, responses to all assessments completed could be centralized. Redundant information could be eliminated by auto-populating fields, such as demographics as well as criminal and prior treatment history. The sharing of information in a centralized hub will also help staff communicate more efficiently, avoid potential errors, and improve placement decisions.

The NDCS should also engage classification and diagnostic staff in the further development of the classification tools. SMEs will no doubt have suggestions for both static and dynamic items to be collected. While not currently validated, these items could be collected, but not scored, with the current classification assessment. Following a one to two-year period of data collection, these additional items could be evaluated for their ability to predict infractions and potentially added to an updated version of the tool. This process has been previously referred to as "item beta testing".

This recommendation would necessitate software resources and programming to assist in electronic submission of the collected data. An additional effort would also be needed to examine the forms and evaluations currently conducted. This would require SMEs to examine assessments and a multitude of paper forms to identify overlaps in content and potential redundancies to be removed.

The timing may be favorable for implementation of this recommendation. With the NDCS's current move to implement the STRONG-R recidivism assessment, a software platform may exist, or be modified to meet the recommended needs. Additionally, the STRONG-R will also likely provide a source of dynamic predictors and collect items that overlap with the newly developed classification tool, where the potential auto-population of criminal history and institutional behaviors will save classification and reclassification assessment labor going forward.

Bed-Space Driven Processes

A repeated theme of the process evaluation was that "classification is a bed-space driven process". The NDCS, like other correctional agencies, will be constantly fighting a battle of resources. That is, overcrowding will likely always be an issue but the degree to which it will impact day-to-day operations will fluctuate. Furthermore, programming needs of the population will change and resources for any and all offender needs are rarely universally available in all facilities. When an influx of inmates creates pressure on the agency to reduce inmate custody level, neither the Hardyman nor the newly created classification tools are designed to remedy an urgent bed space resource demand.

The classification tools constructed were developed to predict infractions, which translates to inmate safety. Unlike the Hardyman tool, the new classification models provide a more specified understanding of violence and serious infraction behavior, while also providing facility/custody level information regarding infraction behavior. However, a classification of "High Violent" does not guarantee that an offender *will* commit a violent infraction, nor more than a classification of "Low Risk" guarantee the offender will not commit misconduct. What the tool can provide is the ability to identify, of those eligible for a custody promotion, which offender is least likely to commit an infraction.

Therefore, an infraction prediction tool is a good start in preparing for the anticipated need to transfer offenders and/or reduce custody levels but should be paired with additional research and information. As indicated in the process evaluation, staff often commented that many needed programs (i.e. substance abuse treatment) were only available in certain facilities, where some offenders required a demotion in custody to receive needed interventions and services. It is therefore recommended that an assessment of NDCS's programming and services be conducted. Taxman's CJ-TRACK simulation tool is an example of how an agency may evaluate its interventions and identify both the quantity and quality of programming needed to address the issues of their population (Taxman, 2013). Similar efforts have been conducted by Hamilton in Washington (Campbell, 2015) and with juvenile populations in Florida (Baglivio, Wolff, Jackowski, & Greenwald, 2015). A statewide waiting list for programming needed, prioritization and sequencing could then be developed to more efficiently manage the needs of offenders. Ultimately, an agency that is the size of NDCS should take inventory of their offenders' needs, the capacity of their evidence-based practices, and make requests for funding to shore up identified gaps.

It is also recommended that a forecasting tool be created. Like a meteorologist monitors weather, forecasts can be made that can help anticipate the ebbs and flows of correctional populations and account for offenders' programming needs. The Washington State Institute for Public Policy (WSIPP) developed such a tool and updates its models with new data annually (Aos, Miller, & Drake, 2006). Working in conjunction with an infraction risk tool, forecasting models such as these will help plan for agency needs and potentially mitigate issues related to overcrowding.

A final consideration regarding bed-space issues concerns overrides. As has been repeatedly described, overrides hinder the efficient use of assessment and classification tools. Assessment tools are designed to create consistency, while overrides are a necessary stop-gap to be used for the exceptional case. A concerted effort was made to integrate as many override factors as possible into the current assessment. However, research regarding when and how often overrides are used should quickly follow the implementation of the new classification models. If it is revealed that overrides are most often used to over-

or under- classify offenders, NDCS could consider eliminating some of the 22 override rationales, alter risk category cut points, or outline new items to collect information used as part of the override decisions.

To summarize, the following recommendations are provided:

1. ***Create an implementation, training, and quality assurance plan.*** This plan should outline how to train current and new staff as well as provide boosters and updates for assessment users. The training should incorporate information around the tools' development and encourage staff to suggest additional modifications to scoring and, practice and policy. The development of training implementation plan should be established and is recommended to include a timeline for a graduated roll out of the new tools, a pilot study to assess system impact and address any modification needs to scoring and cut point placement, a training schedule, and outline for quality assurance checks and future validation analyses.

2. ***Continue improving the tool.*** Adding to the pool of items, particularly dynamic predictors will help improve the current models. Updated versions of the tool can be developed using items collected through current screening and assessment from or through beta tested items developed by staff. In a few years' time, the current tool should be evaluated and updated to improve predictive accuracy.

3. ***Create efficient uses of assessment labor.*** Increases in assessment items collected create additional strains on labor. These additional efforts may be mitigated through a collective effort to eliminate redundant items and streamline data entry and tracking with novel software solutions.

4. ***Create an inventory of interventions and forecast agency needs.*** The infractions tool created is limited in its ability to predict the ebbs and flows of the NDCS population. While bed space restrictions will always be an issue, program inventories, waiting lists, and forecasting efforts will assist in planning and help diminish the pressures related to overcrowding.

5. ***Evaluate override factors and practices.*** Much of the current tool development creation efforts were as a result of the frequency in which both mandatory and discretionary overrides were used in classification and reclassification decisions. To prevent the reoccurrence of this issue and to improve the accuracy and reliability of the tools' application, research is needed to examine when overrides are used and how the current tools can be modified to reduce the need and use of override factors.

6. ***Review of classification policies.*** Related to the use of overrides are NDCS policies that may not reflect current NDCS needs. Based on the recent Justice Reinvestment report (2015), over classification, lack of needed programming, and extended use of overrides was identified as problematic aspects of the NDCS classification system. It is clear from our process evaluation and meetings with SMEs that classification staff are doing their best to maintain safety and attempting to stay within departmental goals and policies. With that said, departmental goals are now conflicting with mandatory overrides and perceived best practices of staff. These policies should be reviewed by the Director's executive and classification staff to determine a consistent approach for classification that removes bottlenecking issues, reduces over classification and overcrowding, and continues to maintain staff, offender, and public safety.

7. ***Pilot testing.*** While the new models are not a completely new procedure for classification staff, they do represent a substantial change to their current processes. It is advisable that a cross-section of classification staff be selected and trained on the new models. These staff should then pilot test the tool examining how the tool functions with the current offender population. This could be done with

new offender assessments or reassessments of initial and reclassifications previously completed. Staff will be encouraged to provide feedback, outlining final model tweaks and improvements to the updated processes. Following pilot testing, a full training should be completed prior to the go-live date. The process of training, testing and full implementation of the tool should take place over the course of one-to-three months, allowing for modifications and adjustments along the way.

Final Considerations

Tool development is an organic process. Staff members need to play key roles as stakeholders in the development, implementation, quality assurance (QA), evaluation, and modifications going forward. Any assessment instrument that is left to languish, unadjusted or without updates, will become obsolete and exist as another task to be completed by staff rather than an informative element of their daily process. NDCS staff members are the users of the tool and need to be utilized to improve its functionality. Their involvement will ultimately increase their confidence and trust in the instrument and more apt to let the results guide day-to-day operations.

Integrating the tool into current risk and needs assessment efforts will provide a more efficient use of staff labor. While there are definite distinctions, the risk of infractions and recidivism are correlated behaviors and the assessment of each should not be siloed. Staff working to reduce negative behavior within the facility and upon reentry will have many common goals. Programs that assist in reducing recidivism will likely impact infraction behavior as well. A future goal of integrating classification and infraction risk prediction with recidivism assessments via software is recommended and will likely improve efficiencies going forward.

Related to the concept of balancing common goals is that of custody designation placement decisions. While the new modifications to the tool will likely reduce the need or overrides, they will not eliminate their use entirely. As there is no prediction model that can be developed to rule out exceptional cases, the guide that a tool's outcome provides should not be absolute. Currently, the newly developed and prior tools provide for a single outcome that can then be overridden from maximum to community custody (and vice versa). While not feasible to implement in the current versions of the tools, it is recommended that the NDCS explore the use of a matrix-style classification system that can incorporate both the risk of infractions and the other stated goals of classification.

Generally, classification systems attempt to first identify offenders with the highest levels of risk and rule them out for placement in lower custody levels. Conversely, and supported by current findings, those offenders with lower levels of risk that are placed in higher custody level increase their risk of infractions and should be ruled out for Max and Medium whenever feasible. However, bed space and other NDCS needs may make it difficult to place an offender in to a single facility type, as indicated by the tool. Furthermore, offender programming and other needs tend to target lower custody designations for delivery of interventions.

Instead, it is beneficial to view risk classification of any tool as a guideline and the resulting outputs (custody designations) along a continuum. Meaning that, although an offender may be categorized as "high violent", they could be deemed eligible for a lower custody if classification staff members determine the offender to be stable and in need of programming offered at a lower custody level. An example matrix is provided in Figure 3. This two-dimensional product identifies both continuums of need and risk, translating corresponding categories in custody designation cells within the table. While criteria of both NDCS and offender needs would need to be operationalized for the vertical axis of the matrix (a task not feasible within the current implementation timeframe), it is recommended that a design such as this be considered going forward. Placing offenders along these continuums offers classification staff flexibility, making decisions based on risk and need simultaneously. The matrix also provides an understanding of graduated elements of both risk and need, providing justifications for placements, rather than overrides of a single determination based only on an offender's risk to infract.

Figure 3. Example Classification Matrix

Need	**Risk**			
	High Violent	High Serious	Moderate	Low
None	Max	Max	Medium	Minimum
Low	Max/Medium	Max/Medium	Minimum	Minimum
Moderate	Max/Medium	Medium	Minimum	Community
Substantial	Medium	Minimum	Community	Community
High	Minimum	Community	Community	Community

Finally, as the NDCS agency, facilities, and offender populations change, so too should the tools that are created. Following implementation, future efforts should be planned to evaluate the newly implemented tools. Newly assessed items are to be evaluated for inclusion in updated versions of the tools. New and more precise methods of selecting and weighting responses will be developed over time and should be implemented where appropriate. As the NDCS continues to take steps forward, utilizing evidence to inform current practices and policies, the data, techniques and accuracy of offender assessments will improve and will need to be consistently evaluated and improved to meet the agency's goals of maintaining inmate, staff, and public safety.

REFERENCES

Aos, S., Miller, M., & Drake, E. (2006). Evidence-Based Public Policy Options to Reduce Future Prison Construction, Criminal Justice Costs, and Crime Rates. Olympia: Washington State Institute for Public Policy.

Austin, J., Coleman, D., Peyton, J., & Johnson, K. D. (2003). *Reliability and validity study of the LSI-R risk assessment instrument*. Washington, DC: Institute on Crime, Justice, and Corrections at the George Washington University.

Baglivio, M., Wolff, K., Jackowski, K., & Greenwald, M. (2015). A Multilevel Examination of Risk/Need Change Scores, Community Context, and Successful Reentry of Committed Juvenile Offenders. *Youth Violence and Juvenile Justice.* 1-24.

Baird, C. S. (2009). *A question of evidence: A critique of risk assessment models used in the justice system*. Madison, WI: National Council on Crime & Delinquency.

Barnoski, R., & Aos, S. (2003). *Washington's Offender Accountability Act: An analysis of the Department of Corrections' risk assessment* (Document No. 03-12-1201). Olympia: Washington State Institute for Public Policy.

Campbell, C. (2015). Evidence-Based Practices Proviso. Spokane: Washington State Institute for Criminal Justice Research: Research in Brief.

Einhorn, H., & Hogarth, R. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, *13*, 171-192.

Hardyman, P. (2005). *Design and Validation of the Nebraska Department of Correctional Services External Classification System*. Criminal Justice Institute. Middletown, CT.

Hamilton, Z., Kigerl, A., Campagna, M., Barnoski, R., Block, L. Lee, S., & van Wormer, J. (2016). Tailoring to Fit: The Development and validation of the STRONG-R Recidivism Risk Assessment. *Criminal Justice and Behavior*. 43(2): 230-263. DOI: 10.1177/0093854815615633.

Irwin, J., & Cressey, D. R. (1962). Thieves, convicts, and the inmate culture. *Social problems, 10*(2), 142-155.

Rice, M., & Harris, G. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, *29*, 615-620.

Silver, E., Smith, W. R., & Banks, S. (2000). Constructing actuarial devices for predicting recidivism: A comparison of methods. *Criminal Justice and Behavior*, *27*, 733-764.

Steyerberg, E. W., Harrell, F. E., Borsboom, M. J. C., Vergouwe, J., & Habbema, D. F. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. Journal of Clinical Epidemiology, 54, 774-781.

Sykes, G. M. (1958). *The society of captives: A study of a maximum security prison*. Princeton, NJ: Princeton University Press.

Taxman, F., The Risk-Need-Responsivity Simulation Tool. Fairfax: The Center for Advancing Correctional Excellence.

Vilinoe, J., Rudes, D., & Taxman, F. (2015). Misalignment In Supervision Implementing Risk/Needs

Assessment Instruments in Probation. *Criminal Justice and Behavior. 1-23.*

Wainer, H. (1976). Estimating coefficients in linear models. *Psychological Bulletin*, *83*, 213-217.

APPENDIX I. Operational Definitions for Study measures

Security threat group - verified: A dichotomous measure, where NDCS staff identify if an offender is a confirmed member of a known gang.

Sentenced for felony: Indicated if the offender's current offense was a felony

Current offense: Categories of the offense for which the offender was incarcerated.

Age at admission: Age at the time of prison admission

Age at first conviction: Recorded age of the offender's first conviction

Highest grade completed: Highest grade at time of admission to prison.

Full time employment/Child care prior: Offender indicated at admission that they possessed full time employment or were the primary child care provider.

Earliest release date: Identified earliest release date of the offender's Earned Discharge Date, parole date, or sentence end date (Tentative Release Date).

Detainer issue: Offender has an identified detainer, hold, or notification

Prior NDCS incarcerations: Count of NDCS incarcerations prior to the current admission date.

Seriousness of instant offense: Seriousness of the offense for which offender was incarcerated.

Prior violent offense: Count of violent offenses prior to admit date, regardless of misdemeanor vs. felony status.

Prior property offense: Count of property offenses prior to admit date, regardless of misdemeanor vs. felony status.

Prior drug offense: Count of drug offenses prior to admit date, regardless of misdemeanor vs. felony status.

Prior sex offense: Count of sex offenses prior to admit date, regardless of misdemeanor vs. felony status.

Prior weapons offense: Count of weapon offenses prior to admit date, regardless of misdemeanor vs. felony status.

Prior escape offense: Count of escape offenses prior to admit date.

Prior domestic violence offense: Count of domestic violence offenses prior to admit date, regardless of misdemeanor vs. felony status.

Prior assault offense: Count of assault offenses prior to admit date, regardless of misdemeanor vs. felony status.

Prior violent-property offense: Count of violent-property offenses prior to admit date, regardless of misdemeanor vs. felony status.

Prior any offense: Ordinal categories counting the number of any offense prior to admit date.

Prior incarceration any infractions: Ordinal measure indicating the number of infractions committed prior to admission date.

Prior incarceration violent infractions: Ordinal measure indicating the number of violent infractions committed prior to admission date.

Prior incarceration serious infractions: Ordinal measure indicating the number of serious (non-violent Class 1 or 2 MRs) infractions committed prior to admission date.

Any Infractions During Current Incarceration: Ordinal measure indicating the number of infractions committed following their admission date.

Violent Infractions During Current Incarceration: Ordinal measure indicating the number of violent infractions committed following their admission date.

Serious Infractions During Current Incarceration: Ordinal measure indicating the number of serious (non-violent Class 1 or 2 MRs) infractions committed following their admission date.

Any Infractions 6 Months Prior to Reclassification: Ordinal measure indicating the number of infractions committed in the six months prior to the reclassification assessment.

Violent Infractions 6 Months Prior to Reclassification: Ordinal measure indicating the number of violent infractions committed in the six months prior to the reclassification assessment.

Serious Infractions 6 Months Prior to Reclassification: Ordinal measure indicating the number of serious (non-violent Class 1 or 2 MRs) infractions committed in the six months prior to the reclassification assessment.

All Prior Any Infractions: Ordinal measure indicating the number of infractions committed the offender's lifetime.

All Prior Violent Infractions: Ordinal measure indicating the number of violent infractions committed the offender's lifetime.

All Prior Serious Infractions: Ordinal measure indicating the number of serious (non-violent Class 1 or 2 MRs) infractions committed the offender's lifetime.

Approved Custody Level: Custody level offender was approved to reside in following the initial classification

Reclass Facility Location: Facility offender was located in at the time of the reclassification assessment.

Facility Location Following Reclass: Facility offender was located in following the reclassification assessment

Transferred to a New Facility Following Reclass: Dichotomous measure indicating if the offender was transferred to a new facility following the reclassification assessment.

Incarceration Duration to Reclass: Ordinal measure of the offender's months/years incarcerated for the current offense to the date of the reclassification assessment.

Non-Serious Infraction Prior to Reclass: Dichotomous measures indicating of 10 or more Class 3 infractions prior to the reclassification assessment

Serious Infractions Prior to Reclass: Dichotomous measures indicating of five or more Class 2 infractions prior to the reclassification assessment

Discretionary overrides: Discretionary override code indicate after initial or reclassification assessment, if any

Mandatory overrides: Mandatory override code indicate after initial or reclassification assessment, if any

Hardyman items: Hardyman assessment closest in days to initial or reclassification assessment date. For more information on these items see Hardyman, 2005.

APPENDIX II. Classification Scoring Forms

Male Initial Classification			
Item	**Violent**	**Serious**	**Non-Serious**
Security Threat Group Verified	54	40	45
Current Offense (not mutually exclusive)			
Violent	1		
Property	11	13	30
Age At Admission - R			
<18	147	135	210
18-19	98	90	140
20-29	49	45	70
30-39	0	0	0
40-49	-49	-45	-70
50-59	-98	-90	-140
60+	-147	-135	-210
Detainer Issue (any)	35	8	
Prior Prison Incarcerations			
0	0	0	
1	1	1	
2	2	2	
3	3	3	
4+	4	4	
Prior Violent Offense			
0	0	0	
1	1	1	
2	2	2	
3	3	3	
4	4	4	
5	5	5	
6	6	6	
7	7	7	
8+	8	8	
Prior Property Offense			
0	0	0	0
1	3	3	4
2	6	6	8
3	9	9	12
4	12	12	16
5	15	15	20
6	18	18	24
7	21	21	28
8+	24	24	32

Prior Sex Offense			
0	0		
1	3		
2	6		
3	9		
4	12		
5+	15		
Prior Weapons Offense			
0	0		0
1	8		7
2+	16		14
Prior Domestic Violence Offense			
0			0
1			2
2+			4
Prior Assault Offense			
0	0		
1	9		
2	18		
3	27		
4	36		
5+	45		
Prior Violent-Property Offense			
0		0	
1		2	
2		4	
3		6	
4		8	
5+		10	
Prior Incarceration Any Infractions			
0		0	0
1-4		1	1
5-9		2	2
10-24		3	3
25+		4	4
Prior Incarceration Violent Infractions			
0	0		0
1	15		14
2+	30		28
Prior Incarceration Serious Infractions			
0		0	
1-5		1	
6-10		2	

11+		3	
Discretionary Overrides			
D3 – Program Participation	3	9	
D4 – NDCS Need	35	32	12
D6 – Protective Custody	84	20	42
Mandatory Overrides			
M3 – ICE Detainer	39		
M4 – Low Severity Detainer			17
M5 – Moderate/High Severity Detainer	1	3	5
Hardyman Items			
1 – Severity of Current Offense			
Highest	21	12	27
High	14	8	18
Moderate	7	4	9
Low	0	0	0
3 – Severity of Prior Convictions			
Highest/High	14	4	12
Moderate	7	2	6
Low/None	0	0	0
6 - Age at First Conviction			
<20	12	36	51
20-27	8	24	34
28-38	4	12	17
39+	0	0	0
8 – Stability Factors			
0 – Neither	0	0	0
1 - Ged/High School Diploma or Employed full time at arrest/Child care	-23	-21	-29
2 - Ged/High School Diploma and Employed full time at arrest/Child care	-46	-42	-58
TOTAL SCORE (ENTER HERE)			
RISK CUT POINTS (IDENTIFY IF COLUMN SCORE EXCEEDS CUT POINT)			
VIOLENT	230		
SERIOUS		200	
NON-SERIOUS			230
SELECT RISK CATEGORY (HIGHEST CATEGORY IF EXCEEDED 1+)			
HIGH VIOLENT	4		

HIGH SERIOUS		3	
MODERATE			2
LOW (DID NOT EXCEED CUT ANY POINT)			1

Female Initial Classification			
Item	**Violent**	**Serious**	**Non-Serious**
Security Threat Group Verified	39	39	69
Sentenced for Felony			3
Current Offense (not mutually exclusive)			
Violent	8		
Property		19	5
Age At Admission - R			
<18	135	132	140
18-19	90	88	70
20-29	45	44	35
30-39	0	0	0
40-49	-45	-44	-35
50-59	-90	-88	-70
60+	-135	-132	-140
Earliest Release Date			
< 6 months	0		
6-12 months	5		
13-18 months	10		
19-24 months	15		
25-36 months	20		
37+ months	25		
Lifer	30		
Detainer Issue (any)		10	
Prior Prison Incarcerations			
0	0	0	
1	20	2	
2	40	4	
3	60	6	
4+	80	8	
Prior Violent Offense			
0	0		
1	3		
2	6		
3	9		
4	12		
5	15		
6	18		
7	21		
8+	24		
Prior Property Offense			

0	0	0	0
1	3	2	2
2	6	4	4
3	9	6	6
4	12	8	8
5	15	10	10
6	18	12	12
7	21	14	14
8+	24	16	16
Prior Weapons Offense			
0	0	0	0
1	17	4	12
2+	34	8	24
Prior Escape Offense			
0	0		0
1	58		19
2+	116		38
Prior Domestic Violence Offense			
0	0		0
1	5		14
2+	10		28
Prior Assault Offense			
0	0		
1	22		
2	44		
3	66		
4	88		
5+	100		
Prior Violent-Property Offense			
0	0		
1	12		
2	24		
3	36		
4	48		
5+	60		
Prior Incarceration Any Infractions			
0		0	0
1-4		1	1
5-9		2	2
10-24		3	3
25+		4	4
Prior Incarceration Violent Infractions			
0	0		0

1	4		44
2+	8		88
Prior Incarceration Serious Infractions			
0		0	
1-5		1	
6-10		2	
11+		3	
D3 – Program Participation		9	9
D4 – NDCS Need	54	32	6
D6 – Protective Custody	151	20	76
Mandatory Overrides			
M4 – Low Severity Detainer			12
M5 – Moderate/High Severity Detainer	55	4	10
Hardyman Items			
1 – Severity of Current Offense			
Highest	45	12	30
High	30	8	20
Moderate	15	4	10
Low	0	0	0
3 – Severity of Prior Convictions			
Highest/High	6	4	30
Moderate	3	2	15
Low/None	0	0	0
4 – Escape History			
Secure w/in 5 Years	16	4	
Secure 5-12 Years	12	3	
Non-secure w/in 3 Years	8	2	
Non-secure 3-7 Years	4	1	
None	0	0	
6 - Age at First Conviction			
<20	57	42	66
20-27	38	28	44
28-38	19	14	22
39+	0	0	0
8 – Stability Factors			
0 – Neither	0	0	0
1 - Ged/High School Diploma or Employed full time at arrest/Child care	-41	-22	-32
2 - Ged/High School Diploma and Employed full time at arrest/Child care	-82	-44	-64
TOTAL SCORE (ENTER HERE)			

RISK CUT POINTS (IDENTIFY IF COLUMN SCORE EXCEEDS CUT POINT)			
VIOLENT	160		
SERIOUS		159	
NON-SERIOUS			50
SELECT RISK CATEGORY (HIGHEST CATEGORY IF EXCEEDED 1+)			
HIGH VIOLENT	4		
HIGH SERIOUS		3	
MODERATE			2
LOW (DID NOT EXCEED CUT ANY POINT)			1

Male Reclassification			
Predictors	**Violent**	**Serious**	**Non-Serious**
Security Threat Group - Verified	58	1	
Sentenced for Felony	1	10	
Current Offense (not mutually exclusive)			
Violent	1		
Property	23	29	
Drug		21	
Age At Reclassification			
<18	174	90	75
18-19	116	60	50
20-29	58	30	25
30-39	0	0	0
40-49	-58	-30	-25
50-59	-116	-60	-50
60+	-174	-90	-75
Highest Grade Completed			
Some College or More			-2
High School Diploma or GED			0
11th Grade or Less			2
Detainer Issue (any)	1		
Prior Prison Incarcerations			
0	0		
1	10		
2	20		
3	30		
4+	40		
Prior Property Offense			
0		0	0
1		4	4
2		8	8
3		12	12
4		16	16
5		20	20
6		24	24
7		28	28
8+		32	32
Prior Drug Offense			
0		0	
1		2	

2		4	
3		6	
4		8	
5		10	
6		12	
7		14	
8+		16	
Prior Sex Offense			
0	0		0
1	8		3
2	16		6
3	24		9
4	32		12
5+	40		15
Prior Weapons Offense			
0	0	0	0
1	4	10	5
2+	8	20	10
Prior Domestic Violence Offense			
0		0	0
1		18	22
2+		36	44
Prior Assault Offense			
0	0	0	0
1	15	3	3
2	30	6	6
3	45	9	9
4	60	12	12
5+	75	15	15
Prior Violent-Property Offense			
0	0		
1	17		
2	34		
3	51		
4	68		
5+	85		
Any Infractions During Current Incarceration			
0-5	0		
6-20	1		
21-50	2		
51-100	3		
101+	4		
Violent Infractions During Current Incarceration			

0	0	0	
1-2	8	24	
3-4	16	36	
5+	24	48	
Violent/ Serious Infractions During Current Incarceration			
0	0		
1-2	23		
3-4	46		
5+	69		
Any Infractions 6 Months Prior to Reclassification			
0	0	0	0
1-4	21	22	24
5-11	42	44	48
12+	63	66	72
Violent Infractions 6 Months Prior to Reclassification			
0	0		
1	7		
2+	14		
All Prior Any Infractions			
0-5			0
6-20			1
21-50			2
51-120			3
121+			4
All Prior Violent Infractions			
0	0		
1-2	18		
3-5	36		
6+	54		
All Prior Serious Infractions			
0		0	
1-2		5	
3-5		10	
6-10		15	
11-25		20	
26-60		25	
61+		30	
Discretionary Overrides			
D3 – Program Participation		1	
D4 – NDCS Need	56	17	
D6 – Protective Custody	39	24	11
D8 – Non-Compliance with Program Rules	89	29	41
Mandatory Overrides			

M3 – ICE Detainer	82		
M4 – Low Severity Detainer		3	18
M5 – Moderate/High Severity Detainer	44		6
M6 – Tentative Release Date 8+ Years	13		
M7 – Tentative Release Date 5+ Years	7		
M9 – *Longer Term Restrictive Housing*		2	
Hardyman Items – R			
1 – Involvement with Drugs or Alcohol			
Two or More	22	20	8
One	11	10	4
None	0	0	0
2 – Escape history			
Secure within 5 years	16	36	
Secure within 5-12 years	12	27	
Non-secure within 3 years	8	18	
Non-secure 3-7 Years	4	9	
None	0	0	
5 – Conviction History			
Highest	36		
High	24		
Moderate	12		
Low/None	0		
7 – Performance in Work/Programming			
Refused program/work or was terminated last 6 mon	12	6	2
Selective compliance/plan, waiting list, or working	6	3	1
Working and compliant with plan	0	0	0
TOTAL SCORE (ENTER HERE)			
RISK CUT POINTS (IDENTIFY IF COLUMN SCORE EXCEEDS CUT POINT)			
VIOLENT	400		
SERIOUS		175	
NON-SERIOUS			110
SELECT RISK CATEGORY (HIGHEST CATEGORY IF EXCEEDED 1+)			
HIGH VIOLENT	4		
HIGH SERIOUS		3	
MODERATE			2

LOW (DID NOT EXCEED CUT ANY POINT)			1

Female Reclassification			
Predictors	**Violent**	**Serious**	**Non-Serious**
Sentenced for Felony	6	38	16
Current Offense (not mutually exclusive)			
Violent	1		
Property	48		
Sex	45		
Age At Reclassification			
<18	72	57	150
18-19	48	38	100
20-29	24	19	50
30-39	0	0	0
40-49	-24	-19	-50
50-59	-48	-38	-100
60+	-72	-57	-150
Highest Grade Completed			
Some College or More	-48	-8	-30
High School Diploma or GED	0	0	0
11th Grade or Less	48	8	30
Earliest Release Date			
< 6 months	0		
6-12 months	11		
13-18 months	22		
19-24 months	33		
25-36 months	44		
37+ months	55		
Detainer Issue (any)		20	
Prior Prison Incarcerations			
0		0	
1		21	
2		42	
3		64	
4+		86	
Prior Violent Offense			
0	0	0	
1	9	4	
2	18	8	
3	36	12	
4	45	16	
5	54	20	

6	63	24	
7	72	28	
8+	4		
Prior Drug Offense			
0		0	
1		1	
2		2	
3		3	
4		4	
5		5	
6		6	
7		7	
8+		8	
Prior Sex Offense			
0		0	
1		4	
2		8	
3		12	
4		16	
5+		20	
Prior Weapons Offense			
0	0	0	0
1	61	3	9
2+	122	6	18
Prior Escape Offense			
0		0	0
1		53	36
2+		106	72
Prior Domestic Violence Offense			
0			0
1			11
2+			22
Prior Assault Offense			
0	0	0	0
1	1	6	2
2	2	12	4
3	3	18	6
4	4	24	8
5+	5	36	10
Any Infractions During Current Incarceration			
0-5	0		
6-20	44		
21-50	88		

51-100	132		
101+	176		
Violent Infractions During Current Incarceration			
0		0	
1-2		8	
3-4		16	
5+		24	
Violent/ Serious Infractions During Current Incarceration			
0	0		
1-2	19		
3-4	38		
5+	57		
Any Infractions 6 Months Prior to Reclassification			
0	0	0	0
1-4	1	9	76
5-11	2	18	152
12+	3	27	228
All Prior Any Infractions			
0-5		0	
6-20		2	
21-50		4	
51-120		6	
121+		8	
All Prior Serious Infractions			
0		0	0
1-2		31	14
3-5		62	28
6-10		94	42
11-25		126	56
26-60		157	70
61+		188	84
Discretionary Overrides			
D3 – Program Participation	34	12	21
D4 – NDCS Need		66	
D8 – Non-Compliance with Program Rules	83	90	66
Mandatory Overrides			
M4 – Low Severity Detainer	72		
M5 – Moderate/High Severity Detainer	85	4	
M7 – Tentative Release Date 5+ Years	75	22	
Hardyman Items – R			
1 – Involvement with Drugs or Alcohol			
Two or More			18

One			9
None			0
5 – Conviction History			
Highest	51		
High	34		
Moderate	17		
Low/None	0		
7 – Performance in Work/Programming			
Refused program/work or was terminated last 6 mon	18	14	26
Selective compliance/plan, waiting list, or working	9	7	13
Working and compliant with plan	0	0	0
TOTAL SCORE (ENTER HERE)			
RISK CUT POINTS (IDENTIFY IF COLUMN SCORE EXCEEDS CUT POINT)			
VIOLENT	305		
SERIOUS		150	
NON-SERIOUS			280
SELECT RISK CATEGORY (HIGHEST CATEGORY IF EXCEEDED 1+)			
HIGH VIOLENT	4		
HIGH SERIOUS		3	
MODERATE			2
LOW (DID NOT EXCEED CUT ANY POINT)			1