



NATIONAL INSTITUTE OF JUSTICE

AUGUST 2024

2023 REVIEW AND REVALIDATION OF THE FIRST STEP ACT RISK ASSESSMENT TOOL



U.S. Department of Justice
Office of Justice Programs
810 Seventh St. N.W.
Washington, DC 20531

Nancy La Vigne, Ph.D.

Director, National Institute of Justice

This and other publications and products of the National Institute of Justice can be found at:

National Institute of Justice

Advancing Justice Through Science

NIJ.ojp.gov

Office of Justice Programs

Building Solutions • Supporting Communities • Advancing Justice

OJP.gov

The National Institute of Justice is the research, development, and evaluation agency of the U.S. Department of Justice. NIJ's mission is to foster and disseminate knowledge and tools derived from objective and rigorous scientific research to inform efforts to promote safety and advance justice.

The National Institute of Justice is a program office of the Office of Justice Programs, which also includes the Bureau of Justice Assistance; the Bureau of Justice Statistics; the Office for Victims of Crime; the Office of Juvenile Justice and Delinquency Prevention; and the Office of Sex Offender Sentencing, Monitoring, Apprehending, Registering, and Tracking.

Opinions or conclusions expressed in this paper are those of the authors and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table of Contents

Executive Summary	3
Part 1: Background	4
Part 2: Method.....	6
Data and Sample	6
Sample Characteristics	7
Analytic Plan.....	10
Part 3: Predictive Validity.....	11
Part 4: Dynamic Validity	14
Part 5: Racial and Ethnic Neutrality	18
Part 6: Discussion and Conclusion	31
References.....	34
Appendix A.....	35
Appendix B.....	37
Appendix C.....	39
Appendix D.....	40

Executive Summary

The First Step Act of 2018 (FSA) mandated the development and implementation of a risk and needs assessment system in the Federal Bureau of Prisons (FBOP). The FSA also required that the U.S. Department of Justice (DOJ) review, validate, and publicly release the risk and needs assessment system on an annual basis. The National Institute of Justice (NIJ) contracted with Dr. Rhys Hester and Dr. Ryan Labrecque as consultants for the annual review and revalidation of the Prisoner Assessment Tool Targeting Estimated Risk and Needs (PATTERN). This document is the fourth review and revalidation report, following USDOJ (2021a), USDOJ (2021b), and USDOJ (2023). The previous review and revalidation reports included FBOP release cohorts from fiscal year (FY) 2009 through FY 2018. The current report analyzes a subsequent cohort of FY 2019 FBOP releasees and evaluates PATTERN for its predictive accuracy, dynamic validity, and racial and ethnic neutrality, as mandated by the FSA. This study analyzes one-, two-, and three-year recidivism outcomes, assesses what proportions of change in risk scores and levels are influenced by the current age item, and provides additional descriptive information on individual items, risk scores and levels, and outcomes by race and ethnic group. Finally, this report provides updates on the actions taken by NIJ and DOJ in the past year and the ongoing efforts to review and improve PATTERN.

The FY 2019 cohort study finds that PATTERN remains a strong and valid predictor of general and violent recidivism at the one-, two-, and three-year follow-up periods, with Area Under the Curve (AUC) statistics ranging from .745 to .776. Comparisons of recidivism rates by risk level category (RLC) and predictive value analyses by risk level grouping also continue to indicate that such risk level designations provide meaningful distinctions of recidivism risk. In addition, individuals can change their risk scores and levels during confinement beyond mere age effects. Those who reduced their RLC from first to last assessment were shown to have the lowest recidivism rates, followed by those who maintained the same risk level and those with a higher risk level, respectively. While the findings continue to indicate PATTERN's accuracy across the five racial and ethnic groups analyzed, there remains evidence that the instruments predict differently across those groups, including overprediction of risk of Black, Hispanic, and Asian males and females, relative to White individuals, on the general recidivism tools.

Part 1: Background

The First Step Act of 2018 (FSA) mandated the development and implementation of a risk and needs assessment system in the Federal Bureau of Prisons (FBOP). The FSA also required the U.S. Department of Justice (DOJ) to review, validate, and publicly release the risk and needs assessment system on an annual basis. The National Institute of Justice (NIJ) contracted with Dr. Rhys Hester (Clemson University) and Dr. Ryan Labrecque (RTI International) to serve as consultants for the annual review and revalidation of the Prisoner Assessment Tool Targeting Estimated Risk and Needs (PATTERN). This document is the fourth review and revalidation report, following USDOJ (2021a), USDOJ (2021b), and USDOJ (2023).¹ The USDOJ (2021a) report documented discrepancies in the scoring, coding, and specification of the PATTERN risk items which precipitated reconstruction of the tool.² The new version, PATTERN version 1.3, was formally adopted in April 2022 (USDOJ 2022a).

The previous USDOJ (2021b) and USDOJ (2023) reports presented revalidation results on the predictive accuracy, dynamic validity, and racial and ethnic neutrality of PATTERN version 1.3. Collectively, these results demonstrated that PATTERN achieved a high degree of predictive accuracy and allowed for a meaningful level of change in risk scores and risk level categories (RLCs) of minimum-, low-, medium-, and high-risk. Although the two prior reports found that PATTERN was a strong predictor of recidivism across the five racial and ethnic groups analyzed, the results also included evidence of differential prediction by race and ethnicity. Among other differences, when controlling for the PATTERN score, Black, Hispanic, and Asian males and females had lower likelihoods of general recidivism compared to white males and females, meaning that PATTERN overpredicted the risk of recidivism for Black, Hispanic, and Asian individuals relative to white individuals (USDOJ 2021b, 39-40). As documented in USDOJ (2021b, 40, fn. 53) those results were not unique to PATTERN version 1.3 but were present in the initial versions of PATTERN as well.³ Thus, transitioning to PATTERN version 1.3 neither exacerbated nor solved these racial bias issues. As reported in USDOJ (2023), DOJ

¹ In addition to these three review and revalidation reports, there were two PATTERN development reports. In July 2019, the U.S. Department of Justice published an initial report describing the development of PATTERN (USDOJ 2019). Following a period of public comment, a second report detailing revisions made to PATTERN (i.e., version 1.2) was released in January 2020 (USDOJ 2020). Under the direction of the attorney general, FBOP began assessing everyone in its custody with PATTERN version 1.2. PATTERN was designed to serve as a risk assessment tool. A separate needs assessment system — the Standardized Prisoner Assessment for Reduction in Criminality (SPARC-13) — has also been developed in compliance with FSA requirements (see FBOP 2022).

² After an independent review of the data, syntax files, and other material used to develop PATTERN, the first review and revalidation report, which identified several discrepancies in the scoring, coding, and specifications of the risk assessment items, was released by NIJ in January 2021. Following this discovery, FBOP updated its risk assessment forms with corrections made to the scoring typos that were published in the USDOJ (2020) report. This version 1.2-revised corrected the typos that were identified in USDOJ (2020) but did not adjust the item weights with the updated data. DOJ elected to postpone updating the item weights until it could discuss and vet the research strategy with the Independent Review Committee and other stakeholders and submit the plan to the attorney general for review and approval. A subsequent report documenting the reconstruction of PATTERN with the corrected data and using the same tool development research methodology (i.e., version 1.3) was published in December 2021.

³ As the report explained, “[T]he differential prediction findings are not the result of the version 1.3 revisions or the changes in data sourcing from FBOP’s automation of PATTERN. The consultants performed the differential prediction analyses on the original 2019 PATTERN 1.2 dataset and obtained substantially similar results for the FY 2014 and FY 2015 validation sample” (USDOJ 2021b, 40, fn. 53).

previously adjusted the risk level category cutpoints which reduced the racial impact of the tool.⁴ NIJ has also considered feedback from subject matter experts and stakeholders and continues to pursue strategies for reducing the differential prediction of PATTERN. Additionally, in September of 2023, FBOP announced a partnership with the Data Science Discovery Program at the University of California, Berkeley, aimed at operationalizing a reconviction (rather than rearrest) recidivism outcome for FBOP release cohorts.⁵ If successful, the reconviction outcome could be used to inform the FSA risk assessment system, which could potentially mitigate some portion of the differential prediction detected across racial and ethnic groups. At this time, however, it is unknown when the reconviction outcomes may be available for analysis. Accordingly, the viability of training and validating PATTERN on reconviction outcomes will not be known for some time.

The FSA mandates the annual review and revalidation of the risk and needs assessment system. The current report documents the revalidation of PATTERN on a subsequent cohort of incarcerated persons — those released from FBOP custody during fiscal year (FY) 2019. As mandated by the FSA, this evaluation focuses its analyses on assessing the predictive accuracy, dynamic validity, and racial and ethnic neutrality of PATTERN. The current study replicates the analyses presented in the prior revalidation report (USDOJ 2023), including by conducting analyses using one-, two-, and three-year post-release recidivism outcomes, assessing to what extent changes in risk scores and levels are influenced by updates to the current age item, and providing additional descriptive information on individual items, risk scores and levels, and outcomes by gender, race, and ethnic group.

⁴ The FSA allows individuals who do not have a disqualifying offense to earn 10 days of time credits for every 30 days of successful participation in evidence-based recidivism reduction programming or productive activities. Individuals determined to be minimum- or low-risk, and whose risk does not increase over two consecutive assessments, earn an additional five days of time credits for every 30 days of successful participation. 28 C.F.R. 523, 541; 18 U.S.C. 3632(d)(4)(A).

⁵ https://www.bop.gov/resources/news/20230927_fbop-partners-with-uc-berkeley.jsp

Part 2: Method

Data and Sample

The sampling frame for this revalidation study was all individuals released from FBOP custody between October 1, 2018, and September 30, 2019 (FY 2019) who did not have an active detainer in place upon release and were not known to have died within three years of release ($N = 42,717$). Individuals were excluded from the study if their initial FBOP classification assessment — the Bureau Risk and Verification Observation (BRAVO) — was not completed under the current version and policy (i.e., FBOP Program Statement on Inmate Security Designation and Custody Classification (PS5100.08); see FBOP 2006; 2019). This criterion was necessary because PATTERN was largely derived from BRAVO and several of the current BRAVO items are needed to score PATTERN. The criterion mostly excluded individuals who were admitted to FBOP custody prior to September 2006.⁶ There were 40,293 individuals who met the study inclusion criteria, including 35,136 males and 5,157 females.

The dataset analyzed in this study was obtained from the FBOP's Office of Research and Evaluation (ORE). The data included demographic, criminal history, and other institutional information necessary to retrospectively calculate each person's PATTERN risk scores at the time of their *first* and *last* BRAVO classifications.⁷ The items used to score PATTERN are described in Appendix A, and the scoring guides for the four version 1.3 instruments are summarized in Appendix B.⁸ The number of items and point structures vary across the four assessments. Within each tool, however, the total risk scores are calculated by summing the values of all included items. These scale scores are then converted into the risk level categories (RLCs) of minimum-, low-, medium-, and high-risk according to the range of values listed in Appendix B.⁹ Those four RLCs were mandated by the FSA. In practice, FBOP further distinguishes between two risk level groupings (RLGs): lower (i.e., minimum- and low-risk RLCs) and higher (i.e., medium- and high-risk RLCs). This distinction is important because although all individuals sentenced for eligible criminal offenses can receive up to 10 days of earned time credit per month toward early release, those in the lower RLG are eligible for an additional five days per month.¹⁰

There are two types of post-release recidivism analyzed in this study. General recidivism is defined as any return to FBOP custody or rearrest within three years of release from FBOP custody, including for driving under the influence and driving while intoxicated but excluding arrests for all other traffic offenses. Violent recidivism is operationalized as any rearrest for an

⁶ There were 2,424 individuals (or 5.7% of the sample) who were excluded due to this criterion.

⁷ By policy, BRAVO assessments are administered upon an individual's arrival to their designated facility, seven months after initial intake, and every 12 months thereafter (see FBOP 2006). Because PATTERN relies on information collected as part of the BRAVO assessment, it must follow BRAVO's schedule of administration for retrospective, pre-FSA assessments.

⁸ For more information on the construction of PATTERN version 1.3, see USDOJ (2021b).

⁹ Note that the attorney general modified the male and female general risk cutpoints in April 2022 (see USDOJ 2022a).

¹⁰ For more information about the earned time credits rule, see FBOP (2022b).

act of violence¹¹ within three years of release from FBOP custody. In addition to three-year outcomes for both variables, this study also includes one- and two-year outcomes.

Sample Characteristics

The descriptive statistics for the study sample are summarized in Table 1, where they are also separated by gender.¹² In both the total and male samples, the largest racial/ethnic group was Black, followed by white, Hispanic, Native American, and Asian. For females, the largest racial/ethnic group was white, followed by Hispanic, Black, Native American, and Asian. The largest age category across all three samples was 30-40 and the smallest age category was > 60. Approximately 12% of the total sample had a Walsh (sex offense) conviction and about a third were convicted of a violent offense. The male sample was approximately 3.3 times more likely than the female sample to have a Walsh conviction, and 3.4 times more likely to have a current conviction for a violent offense. Males also had greater criminal history points and more serious histories of escapes and violence than females. The most common education status category in all three samples was high school degree or GED, and the largest drug program status category was need indicated but no program completion. For both the male and female samples, most individuals did not have a record of a general or serious incident report. Males, however, were more likely than females to have a record of a general or serious incident report both ever and within the last 12 months. Nearly 4% of all three samples were noncompliant with their financial responsibilities to pay victim restitution and support to dependents while in custody. More than half of the total sample completed at least one general program and about a quarter completed at least one work program. The one-, two-, and three-year general recidivism rates were 28.5%, 39.8%, and 46.8% for males and 17.4%, 25.5%, and 31.4% for females. The one-, two-, and three-year violent recidivism rates were 7.8%, 14.0%, and 18.9% for males and 2.0%, 3.9%, and 5.5% for females. The current sample is generally consistent with the prior release cohorts across these numerous metrics. (For comparisons, see USDOJ 2021b, 14-15 and USDOJ 2023, 10-12).

Table 1. Descriptive statistics for the FY 2019 sample, by gender

Measure		% Total (n = 40,293)	% Male (n = 35,136)	% Female (n = 5,157)
Race/ethnicity	<i>White</i>	32.2	30.9	41.0
	<i>Black</i>	39.7	42.0	24.0
	<i>Hispanic</i>	22.3	21.4	28.4
	<i>Asian</i>	1.7	1.7	1.9
	<i>Native American</i>	4.1	4.0	4.8
Current age	<i>> 60</i>	4.9	5.1	4.0
	<i>51-60</i>	11.8	11.9	11.0
	<i>41-50</i>	23.5	23.7	22.5
	<i>30-40</i>	39.6	39.5	39.9
	<i>26-29</i>	12.0	11.8	13.2
	<i>< 26</i>	8.2	8.0	9.5

¹¹ An “act of violence” is defined based on the FBOP Office of Research and Evaluation’s (ORE’s) 19-category offense code classification which mirrors federal code provisions related to violent offenses. See 18 U.S.C. § 16 and 18 U.S.C. § 3559(c)(2)(F). Six classifications are designated as violent: homicide, sexual assault, robbery, assault, weapons, and other violent.; see also FBOP Program Statement 5162.05, Categorization of Offenses.

¹² The descriptives are based on the last PATTERN score administered prior to release.

Walsh with conviction	<i>No</i>	88.2	87.0	96.1
	<i>Yes</i>	11.8	13.0	3.9
Violent offense	<i>No</i>	68.7	65.5	89.9
	<i>Yes</i>	31.3	34.5	10.1
Criminal history points	<i>0-1</i>	27.5	24.8	46.2
	<i>2-3</i>	12.7	12.2	16.8
	<i>4-6</i>	17.5	17.7	16.3
	<i>7-9</i>	14.1	15.0	8.2
	<i>10-12</i>	11.2	11.9	5.8
	<i>13+</i>	17.0	18.5	6.7
History of escapes	<i>None</i>	82.6	81.9	87.9
	<i>> 10 years minor</i>	7.2	7.7	3.5
	<i>5-10 years minor</i>	3.5	3.6	2.8
	<i>< 5 years minor or any serious</i>	6.7	6.9	5.8
History of violence	<i>None</i>	47.1	43.3	73.4
	<i>> 10 years minor</i>	6.5	6.8	4.6
	<i>> 15 years serious</i>	10.2	11.4	2.2
	<i>5-10 years minor</i>	5.3	5.5	4.5
	<i>10-15 years serious</i>	6.8	7.4	1.9
	<i>< 5 years minor</i>	10.3	10.7	7.4
	<i>5-10 years serious</i>	9.1	9.9	3.6
	<i>< 5 years serious</i>	4.8	5.1	2.5
Education status	<i>Not enrolled</i>	20.5	19.7	26.1
	<i>Enrolled in GED</i>	10.3	10.2	10.9
	<i>High school degree/GED</i>	69.1	70.0	63.0
Drug program status	<i>Need indicated/no completion</i>	57.3	57.1	58.5
	<i>Completed non-residential drug treatment</i>	5.9	5.4	5.3
	<i>Completed residential drug treatment</i>	5.5	6.0	6.0
	<i>No need indicated</i>	31.3	31.5	30.2
All incident reports	<i>0</i>	60.8	59.1	72.0
	<i>1</i>	16.0	16.3	14.1
	<i>2</i>	7.8	8.1	5.9
	<i>3+</i>	15.4	16.5	8.0
Serious incident reports	<i>0</i>	74.5	72.7	86.8
	<i>1</i>	13.9	14.6	8.7
	<i>2</i>	5.2	5.6	2.2
	<i>3+</i>	6.4	7.0	2.2
Time since last incident report	<i>12+ months</i>	76.2	75.6	79.8
	<i>7-12 months</i>	3.5	3.6	3.3
	<i>3-6 months</i>	5.0	5.0	4.5
	<i>< 3 months</i>	15.3	15.8	12.4
Time since last serious incident report	<i>12+ months</i>	86.3	85.5	91.5
	<i>7-12 months</i>	2.5	2.6	1.6
	<i>3-6 months</i>	3.4	3.6	2.1
	<i>< 3 months</i>	7.8	8.2	4.8
Financial responsibility refuse	<i>No</i>	96.2	96.2	96.2
	<i>Yes</i>	3.8	3.8	3.8
Programs completed	<i>0</i>	44.8	45.0	43.7
	<i>1</i>	15.7	15.6	15.9
	<i>2-3</i>	14.3	14.0	16.8
	<i>4-10</i>	18.3	18.2	18.5
	<i>11+</i>	6.9	7.2	5.1
Work programs completed	<i>0</i>	76.8	76.9	75.7
	<i>1</i>	14.5	14.2	16.9
	<i>2+</i>	8.7	8.9	7.3

General recidivism	<i>1 year</i>	27.0	28.5	17.4
	<i>2 years</i>	38.0	39.8	25.5
	<i>3 years</i>	44.8	46.8	31.4
Violent recidivism	<i>1 year</i>	7.1	7.8	2.0
	<i>2 years</i>	12.7	14.0	3.9
	<i>3 years</i>	17.2	18.9	5.5

Note: Variable percentages do not all sum to 100.0 due to rounding.

Table 2 reports the total risk scores and RLCs across all four PATTERN instruments at both the first and last assessments.¹³ Across the four tools, the average risk scores were all lower at the last assessment compared to the first. There are also greater proportions of individuals assigned to the minimum- and low-risk RLCs (and fewer in the medium- and high-risk RLCs) at last assessment compared to the first. The distribution of scores and changes in risk levels are similar to those presented in the previous 2018 sample (USDOJ 2023, 12).

Table 2. PATTERN total risk scores and risk level categories, by assessment type

Risk assessment	First Assessment	Last Assessment
Risk level category (range of scores)		
Male general recidivism risk score (SD)	43.0 (20.4)	36.6 (22.3)
Percent minimum (-22 to 5)	3.9	10.2
Percent low (6 to 39)	37.7	41.8
Percent medium (40 to 54)	25.7	24.4
Percent high (55 to 109)	32.6	23.6
Male violent recidivism risk score (SD)	25.4 (12.9)	22.9 (14.1)
Percent minimum (-11 to 7)	8.1	15.5
Percent low (8 to 24)	39.6	38.9
Percent medium (25 to 31)	18.8	16.7
Percent high (32 to 71)	33.5	28.9
Female general recidivism risk score (SD)	26.5 (17.6)	20.3 (20.0)
Percent minimum (-27 to 7)	14.5	28.2
Percent low (8 to 38)	61.2	53.0
Percent medium (39 to 52)	15.6	12.3
Percent high (53 to 102)	8.7	6.5
Female violent recidivism risk score (SD)	4.6 (4.4)	2.4 (5.8)
Percent minimum (-11 to 1)	24.1	46.9
Percent low (2 to 11)	67.8	45.6
Percent medium (12 to 17)	7.7	6.5
Percent high (18 to 30)	0.5	1.1

Note: SD = standard deviation. Percentages do not all sum to 100 due to rounding. Male sample $n = 35,136$; female sample $n = 5,157$.

Additional earned time credit eligibility is determined by classification in the lower RLG on both the general and violent recidivism risk tool. For example, if someone is classified in the lower RLG on the general tool and the higher RLG on the violent tool, then they would be considered higher risk and not eligible for the additional five days of earned time credit for every 30 days of

¹³ As noted in USDOJ (2023), there is a discrepancy in how FBOP operationalizes the education status measure. Individuals are classified as (1) having earned a high school degree or GED, (2) being enrolled in a GED program, or (3) not being enrolled and having no degree earned. As developed (and validated here) the tools considered individuals assigned but not yet enrolled as not participating. FBOP, however, counts those wait listed individuals as participants.

programming completed. As reported in Table 3, 88.5% of males (47.5% + 41.0%) and 86.9% of females (80.3% + 6.6%) were classified in the same RLG on the general and violent tools. The remaining individuals were classified in the higher RLG based on their higher risk designation on one of the two tools.

Table 3. PATTERN overall risk level grouping (RLG) categories, by gender

Males	General lower RLG	General higher RLG
Violent lower RLG	16,673 (47.5%)	2,445 (7.0%)
Violent higher RLG	1,597 (4.5%)	14,421 (41.0%)
Females	General lower RLG	General higher RLG
Violent lower RLG	4,143 (80.3%)	625 (12.1%)
Violent higher RLG	47 (0.9%)	342 (6.6%)

Note: Lower RLG is composed of minimum- and low-risk RLCs, and higher RLG is composed of medium- and high-risk RLCs. Shaded cells indicate agreement between the general and violent tools. Male sample $n = 35,136$; female sample $n = 5,157$.

Analytic Plan

The analytic plan corresponds with the approach taken in the previous USDOJ (2023) report. More specifically, the current review and revalidation focuses on addressing the FSA mandates of predictive validity, dynamic validity, and racial and ethnic neutrality. For predictive validity, the study reports on Area Under the Curve (AUC) statistics, risk level recidivism analyses, and predictive value and false rate analyses (positive and negative predictive values and false positive and negative rates) using one-, two-, and three-year recidivism follow-up periods. For dynamic validity, the study examines changes in risk scores and levels from first to last assessment and conducts additional analyses on what accounts for change in risk scores and levels. Finally, this report provides additional descriptive information on individual items, risk scores and levels, and outcomes by racial and ethnic group. It also examines racial and ethnic neutrality through comparisons of predictive metrics broken out by race and ethnicity, and through differential prediction regression analyses.¹⁴

¹⁴ As in previous reports, these analyses proceeded with a series of four nested logistic regression analyses for each of the four tools, for a total of sixteen regressions. Model 1 included only the categorical race and ethnicity identifier as a predictor of recidivism, with white individuals serving as the reference group. Model 2 included only the PATTERN risk score and assessed whether the score independently predicted recidivism. Model 3 included both the PATTERN risk score and the race and ethnicity identifier. A statistically significant result for a nonwhite group indicated that for a given PATTERN score, members of that group had a different likelihood of recidivism than the white comparison group, on average. A positive result indicated that the nonwhite group was more likely to recidivate compared to the white group (i.e., relative underprediction of risk), and a negative result indicated that the nonwhite group was less likely to recidivate compared to the white group (i.e., relative overprediction of risk). For Model 4, an interaction term between race/ethnicity and the risk score was added to test whether the relationship between race and recidivism varied significantly across changes in the risk score (see also USDOJ 2021b, 37).

Part 3: Predictive Validity

Table 4 presents the results from the AUC analyses that examined the relationship between the total risk scores of the four PATTERN instruments and the recidivism measures at the one-, two-, and three-year follow-up periods.¹⁵ Across all tools and follow-up periods, the AUC values were found to be strong predictors of recidivism (i.e., AUCs $\geq .714$) in all but one comparison. The general male scale was a moderate predictor of recidivism (i.e., AUC $\geq .639$) at the one-year follow-up period. Even the lower bounds of the 95% confidence intervals (CIs) were considered large in terms of predictive strength in 20 of the 24 analyses.¹⁶ The AUCs were decisively higher at last assessment (range = .745 to .776) compared to first (range = .709 to .748).¹⁷ For the general recidivism tools, the AUCs increased with each additional year of follow-up added in both assessment periods. To illustrate, the AUC for the male general recidivism tool at last assessment was .750 at the one-year follow-up period, .767 at the two-year follow-up period, and .775 at the three-year follow-up period. In contrast, the AUCs for the violent recidivism tools were more variable across the different follow-up and assessment periods.

Table 4. PATTERN areas under the curve and 95% confidence intervals, by assessment type and follow-up period

Risk assessment Recidivism follow-up period	AUC [95% CI] First Assessment	AUC [95% CI] Last Assessment
Male general recidivism		
One-year follow-up	.709 [.703, .715]	.750 [.745, .756]
Two-year follow-up	.728 [.723, .734]	.767 [.763, .772]
Three-year follow-up	.739 [.734, .744]	.775 [.771, .780]
Male violent recidivism		
One-year follow-up	.742 [.734, .751]	.767 [.758, .775]
Two-year follow-up	.746 [.739, .752]	.766 [.759, .772]
Three-year follow-up	.748 [.742, .754]	.770 [.765, .776]
Female general recidivism		
One-year follow-up	.733 [.716, .749]	.771 [.755, .787]
Two-year follow-up	.738 [.724, .753]	.772 [.759, .786]
Three-year follow-up	.747 [.734, .761]	.776 [.763, .789]
Female violent recidivism		
One-year follow-up	.718 [.668, .767]	.760 [.716, .804]
Two-year follow-up	.725 [.691, .759]	.751 [.718, .783]
Three-year follow-up	.717 [.689, .746]	.745 [.718, .772]

Note: Male sample $n = 35,136$; female sample $n = 5,157$.

¹⁵ As a supplemental analysis, study results were examined using only the 20,106 males (57.2% of the male sample) and the 4,521 females (87.7% of the female sample) who were statutorily eligible to receive additional earned time credit. These AUC findings, presented in Appendix C, support PATTERN as a strong and valid predictor of recidivism across both assessment types and all three follow-up periods in the general male (range = .705 to .769), violent male (range = .718 to .752), general female (range = .732 to .780), and violent female samples (range = .699 to .755).

¹⁶ This finding should be interpreted cautiously, as three of four 95% CIs had lower bounds that fell below .714 were found in the violent female tool. It should be noted that not only was the female sample size relatively small ($n = 5,157$), but also violent recidivism among females was rare, with only 102 (or 2.0%) events at the one-year follow-up period, 202 (or 3.9%) at the two-year follow-up period, and 285 (or 5.5%) at the three-year follow-up period.

¹⁷ When interpreting these findings, it should be noted that PATTERN was developed using three-year recidivism measures. If the instrument were to be reconstructed using one- or two-year recidivism measures, it is probable that its predictive accuracy in identifying these shorter-term recidivists would be increased.

Figures 1 and 2 display the rates of general and violent recidivism by PATTERN RLC for the male and female samples at the one-, two-, and three-year follow-up periods. These figures provide support for the ability of the RLC designations to effectively distinguish between groups of individuals based on their risk for recidivism. As can be seen in both figures, the rate of recidivism monotonically increases with each successively higher RLC. In the general male recidivism tool, for example, 9.6% of those rated as minimum-risk recidivated during the three-year follow-up period, followed by 31.4% in the low-risk group, 58.9% in the medium-risk group, and 77.5% in the high-risk group.

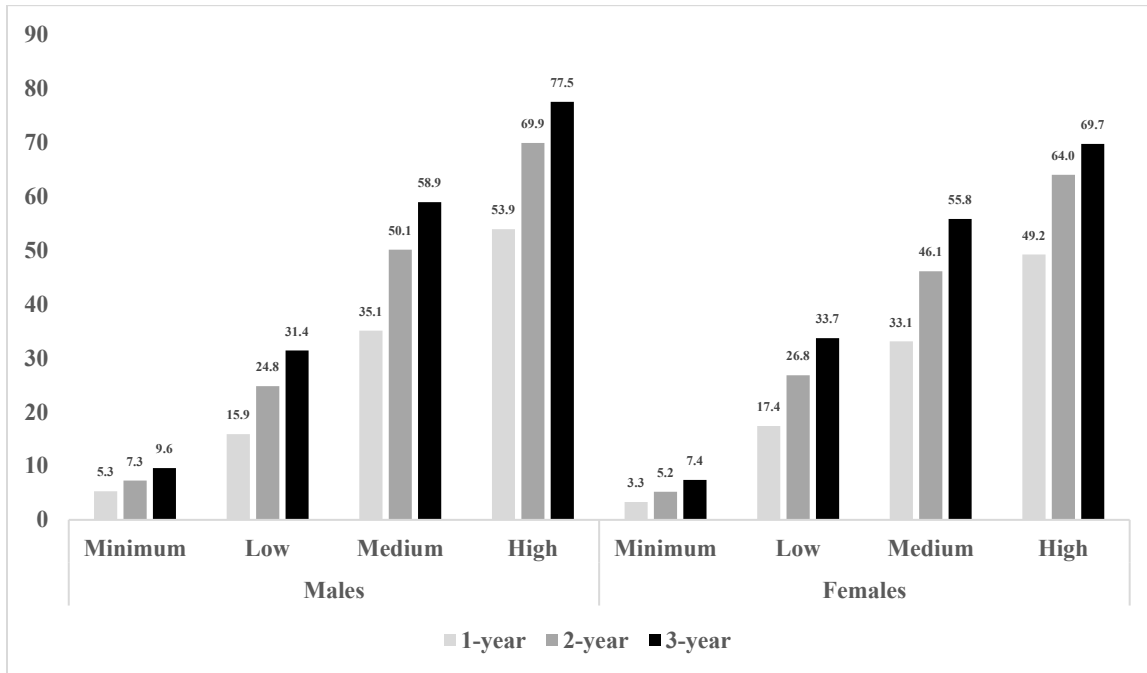


Figure 1. Percentage of males and females with general recidivism outcomes at one-, two-, and three-year follow-up periods by PATTERN last assessment general risk level category

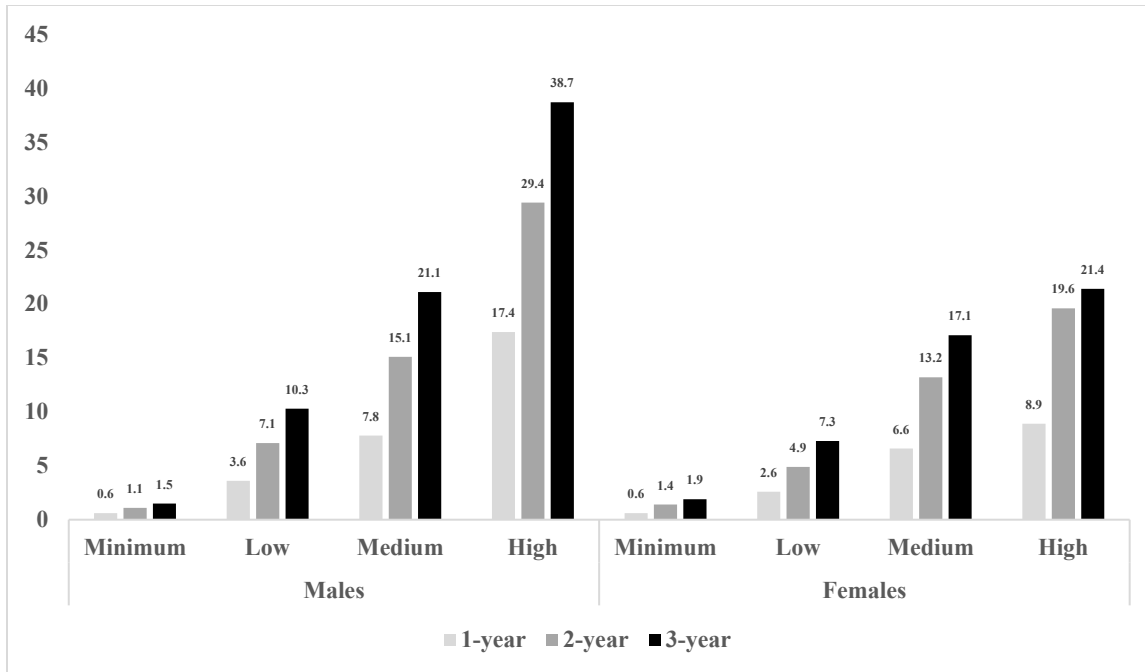


Figure 2. Percentage of males and females with violent recidivism outcomes at one-, two-, and three-year follow-up periods by PATTERN last assessment violent risk level category

Table 5 presents the positive predictive values (PPVs), negative predictive values (NPVs), false positive rates (FPRs), and false negative rates (FNRs) for the three-year recidivism outcomes. While the AUCs rely on the overall PATTERN score, these predictive values are based on the RLGs. The PPV provides a measure of recidivism accuracy in the higher RLG, while the NPV provides a measure of non-recidivism accuracy in the lower RLG.

The higher RLG corresponds to recidivism events (PPVs) for 68% of males and 61% of females on the general tools, and 32% of males and 18% of females on the violent tools. For those classified in the lower RLG on the general tools, 73% of males and 75% of females avoided recidivism (NPVs). For the violent tools, the success rate was 92% for males and 95% for females. The full distribution tables are provided for one-, two-, and three-year outcomes in Appendix D so that interested readers can calculate the predictive values and false rates with the formulas noted in Table 5.

Table 5. Positive predictive values (PPVs), negative predictive values (NPVs), false positive rates (FPRs), and false negative rates (FNRs), by assessment type

Risk assessment	PPV	NPV	FPR	FNR
Male general recidivism	0.68	0.73	0.29	0.30
Male violent recidivism	0.32	0.92	0.38	0.22
Female general recidivism	0.61	0.75	0.11	0.64
Female violent recidivism	0.18	0.95	0.07	0.76

Note: PPV = positive predictive value (proportion of true positives out of all positive predictions); NPV = negative predictive value (proportion of true negatives out of all negative predictions); FPR = false positive rates (proportion of false positives out of all observed non-recidivism); FNR = false negative rate (proportion of false negatives out of all observed recidivism). For these analyses, a higher RLG designation is treated as a positive prediction and a lower RLG designation is treated as a negative prediction. Male sample $n = 35,136$; female sample $n = 5,157$.

Part 4: Dynamic Validity

Table 6 presents changes in PATTERN risk scores and levels from first to last assessment.¹⁸ Across all four instruments, reductions were detected in both the mean scores and RLCs from first to last assessment. For the general recidivism tools, there was 7.5-point reduction detected for males and a 7.7-point reduction for females. In the violent recidivism tools, there was about a 3-point reduction for both males and females. Although most individuals of both genders remained in the same RLC from first to last assessment (between 62% and 68%), approximately 26% to 33% had a lower risk level and 3% to 6% had a higher risk level at their last assessment compared to their first. These findings highlight that individuals can change their risk scores and levels during their period of confinement. They also suggest that while most people remain in the same RLC, a greater proportion of individuals have a lower versus higher score at their final assessment compared to their first.

Table 6. Change in PATTERN risk scores and levels from first to last assessment

Risk assessment	Male sample	Female sample
General recidivism		
Mean change in risk score (SD)	-7.5 (9.5)	-7.7 (8.9)
Percentage at lower risk level	30.5	29.2
Percentage at same risk level	66.5	67.3
Percentage at higher risk level	3.0	3.5
Violent recidivism		
Mean change in risk score (SD)	-3.0 (6.2)	-2.8 (3.7)
Percentage at lower risk level	25.8	33.2
Percentage at same risk level	68.0	61.9
Percentage at higher risk level	6.1	4.9

Note: SD = standard deviation. Percentages do not all sum to 100 due to rounding. Male sample $n = 29,954$; female sample $n = 4,139$.

One of the concerns raised about assessing the dynamic nature of PATTERN is that the age variable, which is weighted heavily and can only be reduced through the passing of time, might be responsible for producing most of the changes in risk scores. To assess the magnitude that changes in age from first to last assessment had on the total change in scores during the same period, we first calculated the average change in score on this item across the four tools. For the male sample, the mean differences in age score from first to last assessment in the general and violent recidivism tools were -2.4 and -1.4 points, respectively. For the female sample, these differences were -1.3 and -0.2 points. Next, we calculated the proportion of the change in age score relative to the total change in score. For the male general recidivism tool, 32.0% of the total change in score was due to change in score on the age variable (i.e., -2.4 points from age item divided by -7.5 points from total score). This same estimate was 46.7% for the male violent tool, 16.9% for the female general tool, and 7.1% for the female violent tool.

Figure 3 illustrates the changes in the overall RLCs for males and females. This figure displays the distribution of the highest RLC assignment on either the general or violent risk tool by gender. As can be seen in the figure, both males and females were more likely to be categorized

¹⁸ The dynamic analyses included only the individuals with more than one assessment available. More specifically, there were 29,954 males (or 85.3% of the male sample) and 4,139 females (or 80.3% of the female sample) who had at least two assessments completed during their period of confinement.

in the lower RLG at last assessment relative to first. More specifically, 9.5% of males and 6.8% of females were less likely to be identified as minimum- or low-risk at the last assessment compared to first. It should be noted that while the difference for females is lower than that for males, this is influenced by the fact that a higher proportion of females were already in the lower RLG at first assessment compared to males (75.2% versus 40.4%, respectively). The proportion of females in the minimum-risk classification, however, more than doubled between first and last assessment (from 15.1% to 32.2%).

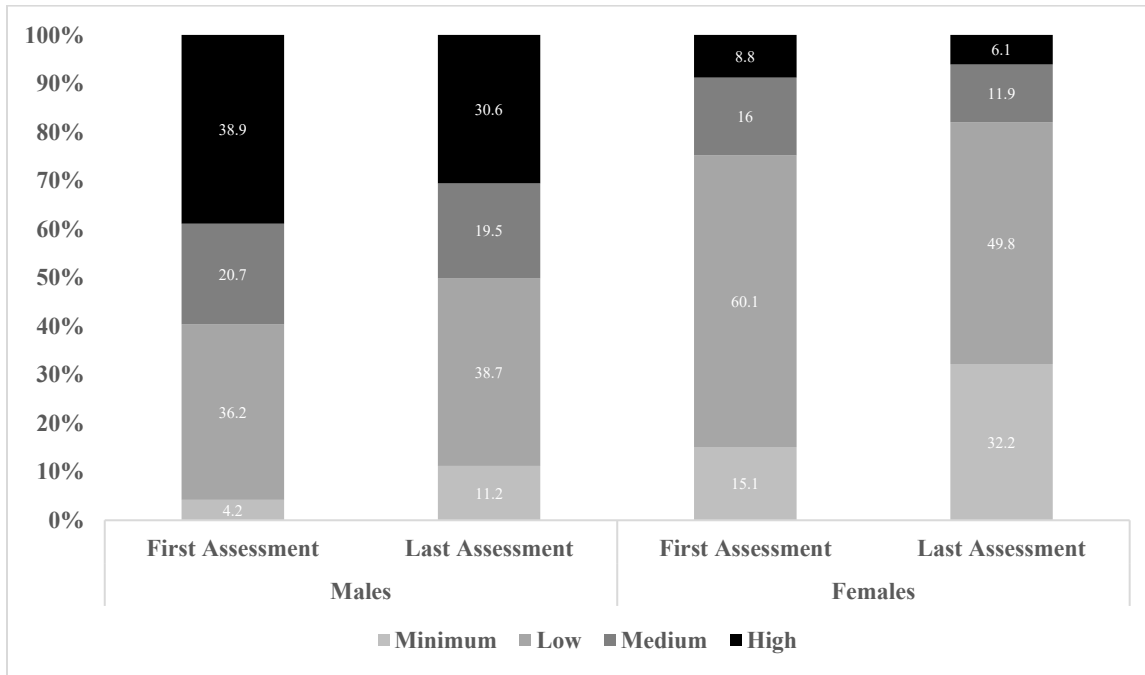


Figure 3. Percentage of males and females assigned to each of the overall PATTERN risk level categories at first and last assessment

Table 7 assesses how changes in RLC relate to rates of recidivism across the one-, two-, and three-year follow-up periods. In this assessment, individuals are grouped into one of three categories: those with a *lower*, the *same*, or a *higher* RLC at last assessment compared to first. Across the four instruments and three follow-up periods, individuals with a lower risk level at last assessment demonstrated the lowest recidivism rates, followed by those with the same risk level and those with a higher risk level in all comparisons with one exception. Females categorized in the same violent risk level at first and last assessment had the same percentage of recidivism as those with a higher risk level at last compared to first (2%). In totality, these findings emphasize that individuals who lower their risk score are at a reduced risk of recidivating in the community upon release, whereas those who elevate their risk score are at a greater risk of recidivating.¹⁹

¹⁹ The violent female findings should all be interpreted cautiously as there were only 202 females who had a higher violent risk level at last assessment compared to first assessment. Among those in the higher risk level category, there were only 4 violent recidivists in year 1, 16 in year 2, and 23 in year 3.

Table 7. Percentage of individuals who recidivated by change in PATTERN risk level status from first to last assessment, by follow-up period

Risk assessment	One-year follow-up (%)	Two-year follow-up (%)	Three-year follow-up (%)
Male General Recidivism			
Lower risk level	17.9	28.1	35.5
Same risk level	29.4	40.6	47.4
Higher risk level	42.6	57.5	64.8
Male Violent Recidivism			
Lower risk level	3.7	7.5	10.4
Same risk level	8.5	15.1	20.2
Higher risk level	9.0	16.2	23.5
Female General Recidivism			
Lower risk level	10.8	16.6	22.7
Same risk level	15.8	23.6	29.1
Higher risk level	35.0	47.6	55.9
Female Violent Recidivism			
Lower risk level	0.9	2.4	3.4
Same risk level	2.0	3.6	5.2
Higher risk level	2.0	7.9	11.4

Note: Lower risk level = lower risk level assigned at last assessment compared to first assessment; same risk level = same risk level assigned at last assessment as at first assessment; higher risk level = higher risk level assigned at last assessment compared to first assessment. Male sample $n = 29,954$; Female sample $n = 4,139$.

Table 8 presents the results of the logistic regression analyses of the first assessment scores and changes in scores from first to last assessment predicting recidivism at the one-, two-, and three-year follow-up periods across the four PATTERN tools. In all 12 of the models, both the first assessment score and the change in risk score were predictive of recidivism ($p < .001$). When holding the first assessment score constant, for every 1-point increase in the total general and violent male scores from first to last assessment, there was a corresponding 6% and 7% increase in the odds of general and violent rearrest, respectively. For the female models, there was similarly a 6% to 7% increase in the odds of general recidivism and a 12% to 13% increase in the odds of violent recidivism. These results confirm that increases in PATTERN risk scores during incarceration are associated with higher levels of recidivism and decreases in scores are associated with lower levels of recidivism.

Table 8. Logistic regression of first PATTERN assessment score and change in score from first to last assessment predicting recidivism at the one-, two-, and three-year follow-up periods

Risk assessment	One-year follow-up	Two-year follow-up	Three-year follow-up
Male General Recidivism			
First assessment score	1.05	1.05	1.05
Change in risk score	1.06	1.06	1.06
Constant	0.07	0.10	0.13
Model χ^2	5,078.3	6,861.0	7,604.2
Nagelkerke R^2	.228	.279	.300
Male Violent Recidivism			
First assessment score	1.08	1.08	1.09
Change in risk score	1.07	1.07	1.07
Constant	0.01	0.02	0.02
Model χ^2	2,076.3	3,391.1	4,508.2
Nagelkerke R^2	.164	.198	.229
Female General Recidivism			

First assessment score	1.05	1.05	1.06
Change in risk score	1.07	1.07	1.06
Constant	0.05	0.09	0.12
Model χ^2	570.3	742.2	846.4
Nagelkerke R^2	.226	.251	.266
Female Violent Recidivism			
First assessment score	1.17	1.18	1.18
Change in risk score	1.13	1.12	1.12
Constant	0.01	0.02	0.02
Model χ^2	60.3	127.6	161.4
Nagelkerke R^2	.094	.118	.118

Note: Reported values are odds ratios. All findings are statistically significant at the .001 level. Male sample $n = 29,954$; Female sample $n = 4,139$.

Part 5: Racial and Ethnic Neutrality

Tables 9 and 10 provide the PATTERN item scores and recidivism outcomes for males and females by race and ethnic group. There are numerous differences found in the item scores across the five racial and ethnic groups in both genders. The white and Asian individuals tend to be older than the Black, Hispanic, and Native American individuals. For example, 28.6% of white males and 21.3% of Asian males are age 51 or older compared to 12.0% of Black males, 10.8% of Hispanic males, and 10.1% of Native American males. Similar but less pronounced age differences appear for the females; for example, 19.1% of white females and 32.3% of Asian females are age 51 or older compared to 14.7% of Black females, 9.4% of Hispanic females, and 6.4% of Native American females. More than 30% of white, Hispanic, and Asian males fall into the lowest criminal history point category (0-1), while only 12.1% of Black males and 21.7% of Native American males fall into that category. While Hispanic and Asian females were also the most likely to fall into the lowest criminal history point category (57.7% and 66.7%, respectively), white, Black, and Native American females were much less likely to fall in that category (37.7%, 48.5%, and 30.6%, respectively). Native American and Black individuals from both genders are also more likely to have a violent conviction than white, Hispanic, and Asian individuals. There are also some measures with group similarities, including program completion and work program completion among white, Black, Hispanic, and Asian males and females.

There were meaningful differences in recidivism rates detected across the racial and ethnic groups of both genders. Black and Native American males had the highest rates of observed three-year general recidivism (50.0% and 72.4%) and violent recidivism (24.8% and 27.5%), while Asian males had the lowest rates of these two outcomes (27.7% and 8.8%). The female recidivism rates exhibited a notably different trend: white and Native American females had the highest rates of observed three-year general recidivism (34.2% and 54.8%), and Black and Native American females had the highest rates of observed three-year violent recidivism (7.0% and 11.3%). Asian females had the lowest rates of both general and violent recidivism (16.7% and 1.0%).²⁰ It is also important to point out that Black and Hispanic females both had lower rates of general recidivism (26.6% and 28.2%) than white females (34.2%).

²⁰ Out of the 96 Asian females, only one had a new violent arrest and it occurred in the first year.

Table 9. Descriptive statistics for the FY 2019 male release sample, by race/ethnicity

Measure	% White (n = 10,869)	% Black (n = 14,750)	% Hispanic (n = 7,514)	% Asian (n = 582)	% Native American (n = 1,421)	
Current age	> 60	10.2	2.7	2.6	6.7	2.5
	51-60	18.4	9.3	8.2	14.6	7.6
	41-50	26.7	24.3	19.0	25.8	17.9
	30-40	34.3	42.6	40.3	39.9	43.5
	26-29	7.2	13.1	15.3	8.8	16.5
	< 26	3.2	7.9	14.6	4.3	12.0
Walsh with conviction	No	78.6	91.0	94.3	90.0	70.0
	Yes	21.4	9.0	5.7	10.0	30.0
Violent offense	No	72.3	56.6	79.0	78.9	30.5
	Yes	27.7	43.4	21.0	21.1	69.5
Criminal history points	0-1	33.2	12.1	35.9	51.4	21.7
	2-3	12.1	9.7	15.5	12.4	20.1
	4-6	15.3	19.5	17.3	13.7	20.3
	7-9	11.6	18.8	12.3	8.1	17.5
	10-12	10.0	15.8	7.9	4.8	10.3
	13+	17.7	24.0	11.0	9.6	9.9
History of escapes	None	83.3	79.4	87.9	88.7	61.4
	> 10 years minor	7.5	9.5	4.0	4.1	11.2
	5-10 years minor	3.2	4.2	2.4	2.4	7.4
	< 5 years minor or any serious	6.0	6.9	5.7	4.8	20.1
History of violence	None	52.5	31.8	55.5	67.0	17.8
	> 10 years minor	6.6	8.4	4.7	5.2	3.6
	> 15 years serious	9.5	15.5	7.1	6.7	7.4
	5-10 years minor	5.1	6.4	4.6	2.6	5.1
	10-15 years serious	6.2	9.3	5.8	4.5	7.9
	< 5 years minor	9.2	11.6	10.6	6.0	15.3
	5-10 years serious	7.7	11.6	7.6	4.5	22.0
	< 5 years serious	3.2	5.4	4.1	3.6	21.0
Education status	Not enrolled	12.0	19.6	30.0	20.3	25.3
	Enrolled in GED	4.5	12.6	13.8	7.7	11.7
	High school degree/GED	83.5	67.8	56.2	72.0	63.1
Drug program status	Need indicated/no completion	52.6	56.7	61.0	45.0	81.6
	Completed non-residential drug treatment	5.8	6.4	5.7	5.0	3.9
	Completed residential drug treatment	6.1	5.1	5.7	5.7	2.0
	No need indicated	35.5	31.8	27.6	44.3	12.6

All incident reports	<i>0</i>	64.5	54.0	59.6	69.1	64.3
	<i>1</i>	15.8	17.0	16.1	17.2	13.2
	<i>2</i>	6.8	9.1	8.4	5.3	7.0
	<i>3+</i>	13.0	19.8	15.9	8.4	15.6
Serious incident reports	<i>0</i>	77.9	68.4	72.4	84.7	74.4
	<i>1</i>	11.9	16.8	14.9	9.3	13.0
	<i>2</i>	4.1	6.7	6.0	3.3	5.1
	<i>3+</i>	6.0	8.2	6.6	2.7	7.5
Time since last incident report	<i>12+ months</i>	79.0	74.1	74.0	82.8	71.6
	<i>7-12 months</i>	3.0	3.7	4.1	3.3	3.2
	<i>3-6 months</i>	4.4	5.1	5.6	2.9	6.3
	<i>< 3 months</i>	13.6	17.0	16.3	11.0	18.8
Time since last serious incident report	<i>12+ months</i>	87.7	84.8	84.3	92.4	80.6
	<i>7-12 months</i>	2.3	2.8	2.6	2.1	2.6
	<i>3-6 months</i>	3.1	3.6	4.2	1.2	5.0
	<i>< 3 months</i>	6.9	8.7	8.9	4.3	11.8
Financial responsibility refuse	<i>No</i>	97.4	95.8	95.4	99.0	93.0
	<i>Yes</i>	2.6	4.2	4.6	1.0	7.0
Programs completed	<i>0</i>	44.2	42.5	47.7	42.4	64.0
	<i>1</i>	16.8	15.4	14.9	13.6	13.9
	<i>2-3</i>	14.0	14.4	14.0	13.9	9.3
	<i>4-10</i>	17.5	20.0	17.2	21.5	9.1
	<i>11+</i>	7.5	7.6	6.3	8.6	3.7
Work programs completed	<i>0</i>	77.0	74.3	79.4	81.1	88.2
	<i>1</i>	14.7	15.0	13.3	12.9	7.9
	<i>2+</i>	8.3	10.7	7.3	6.0	3.9
General recidivism	<i>1 year</i>	27.2	29.2	25.4	15.5	52.4
	<i>2 years</i>	37.2	42.1	35.5	23.4	65.3
	<i>3 years</i>	43.2	50.0	42.3	27.7	72.4
Violent recidivism	<i>1 year</i>	5.3	10.5	5.8	3.6	11.7
	<i>2 years</i>	9.6	18.5	10.6	7.4	20.1
	<i>3 years</i>	13.0	24.8	15.2	8.8	27.5

Note: Variable percentages do not all sum to 100.0 due to rounding.

Table 10. Descriptive statistics for the FY 2019 female release sample, by race/ethnicity

Measure	% White (n = 2,112)	% Black (n = 1,237)	% Hispanic (n = 1,464)	% Asian (n = 96)	% Native American (n = 248)	
Current age	> 60	5.5	3.1	2.4	9.4	2.4
	51-60	13.6	11.6	7.0	22.9	4.0
	41-50	24.3	23.9	18.5	27.1	21.0
	30-40	40.2	39.6	38.9	30.2	48.0
	26-29	10.6	12.6	17.4	9.4	14.1
	< 26	5.7	9.1	15.8	1.0	10.5
Walsh with conviction	No	95.8	94.3	98.1	92.7	96.4
	Yes	4.2	5.7	1.9	7.3	3.6
Violent offense	No	91.2	85.6	95.6	91.7	64.9
	Yes	8.8	14.4	4.4	8.3	35.1
Criminal history points	0-1	37.7	48.5	57.7	66.7	30.6
	2-3	17.9	14.3	16.2	10.4	25.4
	4-6	19.4	14.9	12.0	9.4	24.2
	7-9	9.7	7.8	6.1	4.2	11.3
	10-12	7.0	6.2	4.2	3.1	5.2
	13+	8.3	8.3	3.8	6.3	3.2
History of escapes	None	86.5	88.7	90.9	89.6	78.2
	> 10 years minor	4.7	3.4	2.0	0.0	2.8
	5-10 years minor	3.1	2.5	2.5	2.1	4.4
	< 5 years minor or any serious	5.7	5.4	4.6	8.3	14.5
History of violence	None	77.3	65.2	78.5	87.5	45.6
	> 10 years minor	4.9	5.4	3.3	1.0	7.3
	> 15 years serious	2.2	2.7	1.7	1.0	2.0
	5-10 years minor	4.0	5.5	3.8	3.1	9.3
	10-15 years serious	1.4	3.4	1.2	1.0	2.8
	< 5 years minor	6.1	9.3	6.8	2.1	14.1
	5-10 years serious	2.7	5.0	2.7	1.0	9.7
	< 5 years serious	1.6	3.5	1.9	3.1	9.3
Education status	Not enrolled	19.5	24.8	35.7	21.9	34.7
	Enrolled in GED	6.4	11.8	16.7	8.3	10.9
	High school degree/GED	74.1	63.4	47.6	69.8	54.4
Drug program status	Need indicated/no completion	62.2	52.1	56.3	40.6	78.6
	Completed non-residential drug treatment	6.4	4.8	4.4	4.2	4.8
	Completed residential drug treatment	8.1	4.6	4.1	10.4	4.8
	No need indicated	23.2	38.6	35.2	44.8	11.7

All incident reports	<i>0</i>	75.5	65.8	71.4	79.2	73.4
	<i>1</i>	13.2	15.4	14.3	14.6	14.5
	<i>2</i>	5.4	7.2	6.0	2.1	4.4
	<i>3+</i>	5.9	11.6	8.3	4.2	7.7
Serious incident reports	<i>0</i>	88.4	83.3	87.2	90.6	87.1
	<i>1</i>	8.1	9.5	9.0	7.3	8.5
	<i>2</i>	1.8	3.5	1.9	1.0	2.8
	<i>3+</i>	1.7	3.8	1.9	1.0	1.6
Time since last incident report	<i>12+ months</i>	82.9	75.4	78.8	86.5	79.4
	<i>7-12 months</i>	3.1	3.9	3.3	1.0	2.0
	<i>3-6 months</i>	4.2	4.6	5.0	0.0	4.4
	<i>< 3 months</i>	9.8	16.1	13.0	12.5	14.1
Time since last serious incident report	<i>12+ months</i>	92.8	90.0	90.9	95.8	90.7
	<i>7-12 months</i>	1.4	2.6	1.4	0.0	1.2
	<i>3-6 months</i>	1.9	2.3	2.3	0.0	1.6
	<i>< 3 months</i>	3.9	5.1	5.4	4.2	6.5
Financial responsibility refuse	<i>No</i>	97.0	94.6	96.4	97.9	95.6
	<i>Yes</i>	3.0	5.4	3.6	2.1	4.4
Programs completed	<i>0</i>	42.2	40.8	45.6	37.5	62.1
	<i>1</i>	17.4	16.8	13.1	17.7	14.1
	<i>2-3</i>	16.2	17.5	18.0	19.8	10.5
	<i>4-10</i>	19.1	18.4	19.0	19.8	10.1
	<i>11+</i>	5.1	6.5	4.2	5.2	3.2
Work programs completed	<i>0</i>	75.1	78.7	74.4	74.0	75.0
	<i>1</i>	16.6	14.6	19.0	17.7	18.5
	<i>2+</i>	8.3	6.7	6.6	8.3	6.5
General recidivism	<i>1 year</i>	19.2	14.4	14.8	14.6	33.9
	<i>2 years</i>	27.4	21.1	23.2	15.6	48.4
	<i>3 years</i>	34.2	26.6	28.2	16.7	54.8
Violent recidivism	<i>1 year</i>	1.8	1.9	2.1	1.0	3.2
	<i>2 years</i>	3.1	4.5	4.2	1.0	7.3
	<i>3 years</i>	4.4	7.0	5.3	1.0	11.3

Note: Variable percentages do not all sum to 100.0 due to rounding.

Table 11 summarizes the risk score and RLC distributions by race and ethnicity. There are clear differences in these outcomes across the five racial and ethnic groups of both genders, as would be expected given the group-based differences in the risk items above. For example, Black and Native American males had higher average general recidivism risk scores at final assessment (42.8 and 45.6 points) than white, Hispanic, and Asian males (29.8, 33.7, and 22.9 points). Black and Native American males were also more likely to be classified in the higher RLG of the general recidivism tool (59.1% and 65.3%) than white, Hispanic, and Asian males (38.0%, 39.2%, and 26.0%).²¹ Native American females had much higher general recidivism scores at final assessment (28.6 points) than white, Black, Hispanic, and Asian females, who all had similar general recidivism scores (20.3, 20.5, 19.5, and 9.6 points). Native American females were also more likely to be classified in the higher RLG of the general recidivism tool (21.3%) than white, Black, Hispanic, and Asian females (5.4%, 11.0%, 5.6%, and 5.2%).

²¹ These higher RLG values are based on the sum of “percent medium” and “percent high” in Table 11.

Table 11. PATTERN total risk scores and risk level categories, by assessment type and race/ethnicity

Risk assessment	White		Black		Hispanic		Asian		Native American	
	First	Last	First	Last	First	Last	First	Last	First	Last
Male general recidivism										
Mean risk score (SD)	36.2 (22.3)	29.8 (24.5)	50.0 (17.7)	42.8 (19.6)	38.9 (18.1)	33.7 (20.4)	29.6 (21.2)	22.9 (22.8)	48.0 (16.5)	45.6 (18.7)
Percent minimum	9.0	19.8	1.1	3.9	2.5	8.9	9.6	25.1	0.6	2.4
Percent low	45.5	42.2	25.5	37.0	50.8	51.9	59.8	49.0	27.2	32.3
Percent medium	20.8	19.2	29.0	29.0	25.4	22.5	14.3	12.9	35.8	31.9
Percent high	24.8	18.8	44.4	30.1	21.2	16.7	16.3	13.1	36.4	33.4
Male violent recidivism										
Mean risk score (SD)	20.9 (13.5)	18.2 (14.9)	29.6 (11.6)	26.6 (12.7)	22.8 (11.3)	21.1 (12.7)	17.5 (12.7)	14.6 (13.8)	32.6 (10.9)	32.3 (12.3)
Percent minimum	16.7	28.2	2.7	6.8	6.5	14.9	22.2	37.1	1.5	3.5
Percent low	43.6	37.3	30.8	37.3	53.4	47.3	49.7	38.5	22.2	23.1
Percent medium	15.6	13.3	22.1	19.9	17.3	16.1	11.7	10.1	18.9	15.7
Percent high	24.1	21.2	44.4	35.9	22.8	21.7	16.5	14.3	57.4	57.7
Female general recidivism										
Mean risk score (SD)	27.4 (18.3)	20.3 (20.3)	26.5 (18.3)	20.5 (21.0)	24.8 (15.9)	19.5 (18.4)	16.9 (19.6)	9.6 (21.1)	32.1 (15.0)	28.6 (17.4)
Percent minimum	16.1	28.6	14.8	31.1	12.0	26.5	37.5	56.3	4.8	10.1
Percent low	56.6	51.8	59.2	47.9	69.5	58.4	45.8	32.3	67.7	65.3
Percent medium	17.4	13.5	15.8	12.4	12.9	10.2	7.3	7.3	18.1	15.3
Percent high	9.9	6.1	10.3	8.6	5.5	4.8	9.4	4.2	9.3	9.3
Female violent recidivism										
Mean risk score (SD)	4.6 (4.1)	2.0 (5.4)	4.9 (4.9)	2.8 (6.3)	4.0 (3.9)	2.2 (5.4)	2.6 (4.7)	-0.2 (5.8)	7.7 (4.7)	6.5 (5.9)
Percent minimum	21.2	47.3	27.5	47.4	26.8	49.1	47.9	67.7	6.9	19.4
Percent low	72.6	47.3	60.5	41.6	68.1	45.3	44.8	27.1	69.8	59.3
Percent medium	6.0	4.9	11.1	8.9	4.9	5.0	7.3	5.2	21.0	16.9
Percent high	0.2	0.5	0.9	2.1	0.2	0.6	0.0	0.0	2.4	4.4

Note: SD = standard deviation. Percentages do not all sum to 100 due to rounding. White male sample $n = 10,869$; Black male sample $n = 14,750$; Hispanic male sample $n = 7,514$; Asian male sample $n = 582$; Native American male sample $n = 1,421$; white female sample $n = 2,112$; Black female sample $n = 1,237$; Hispanic female sample $n = 1,464$; Asian female sample $n = 96$; Native American female sample $n = 248$.

Table 12 reports the results of the AUC analyses between the PATTERN total risk scores and the recidivism measures at the one-, two-, and three-year follow-up periods by race and ethnicity. Across the four tools, five racial and ethnic groups, and three follow-up periods, the AUC statistics indicate that the PATTERN tools were strong predictors of recidivism (i.e., AUCs \geq .714) in 91 of the 120 analyses (or 75.8%). In addition, 24 of the 120 analyses (or 20.0%) were considered moderate predictors of recidivism (i.e., AUCs \geq .639).²² There was, however, a wide range of AUCs detected across the racial and ethnic groups. For example, the AUCs from the last assessment and three-year follow-up period ranged from a low of .626 (Native American female violent recidivism) to a high of .860 (Asian female general recidivism).²³ Additionally, the AUCs were higher at last assessment compared to the first in all but one instance (Hispanic female two- year violent recidivism). These results are consistent with those reported on the prior 2018 release cohort (see USDOJ 2023, 28).

²² The poor predictive validity of the violent recidivism scale among Asian and Native American females should be interpreted cautiously as these estimates are based on small sample sizes ($n = 96$ and 248 , respectively). Additionally, there was only one of the 96 Asian females who had a violent recidivism event across all three observation years.

²³ As noted earlier, however, these findings should be interpreted cautiously, as the Asian and Native American female sample sizes are small ($n = 88$ and 287 , respectively).

Table 12. PATTERN AUCs and 95% confidence intervals, by assessment type and race/ethnicity

Risk assessment	One-year follow-up		Two-year follow-up		Three-year follow-up	
	First	Last	First	Last	First	Last
Male general recidivism						
White	.743 [.733, .753]	.772 [.762, .782]	.765 [.756, .774]	.792 [.783, .800]	.776 [.767, .784]	.799 [.791, .808]
Black	.681 [.672, .690]	.734 [.725, .742]	.695 [.687, .704]	.746 [.738, .754]	.704 [.696, .712]	.753 [.745, .760]
Hispanic	.721 [.709, .734]	.754 [.741, .766]	.735 [.724, .746]	.766 [.755, .777]	.739 [.727, .750]	.768 [.758, .779]
Asian	.820 [.777, .862]	.831 [.788, .874]	.844 [.809, .878]	.853 [.818, .889]	.847 [.813, .880]	.860 [.826, .893]
Native American	.670 [.642, .697]	.694 [.667, .722]	.697 [.668, .726]	.719 [.690, .747]	.711 [.681, .741]	.734 [.704, .764]
Male violent recidivism						
White	.769 [.752, .786]	.786 [.769, .802]	.774 [.761, .787]	.789 [.776, .802]	.774 [.762, .786]	.790 [.779, .801]
Black	.698 [.685, .711]	.734 [.722, .747]	.702 [.692, .713]	.736 [.726, .746]	.706 [.697, .715]	.744 [.735, .752]
Hispanic	.748 [.725, .771]	.772 [.749, .794]	.739 [.722, .756]	.753 [.736, .769]	.743 [.728, .757]	.760 [.746, .774]
Asian	.805 [.713, .896]	.814 [.735, .893]	.832 [.775, .889]	.855 [.807, .904]	.809 [.753, .866]	.829 [.777, .882]
Native American	.685 [.645, .726]	.712 [.672, .751]	.708 [.676, .740]	.722 [.690, .753]	.687 [.657, .717]	.708 [.679, .737]
Female general recidivism						
White	.709 [.684, .734]	.753 [.730, .777]	.717 [.695, .740]	.762 [.741, .783]	.739 [.718, .760]	.775 [.755, .795]
Black	.768 [.732, .803]	.803 [.770, .835]	.762 [.731, .793]	.794 [.765, .823]	.757 [.728, .786]	.783 [.755, .810]
Hispanic	.724 [.688, .760]	.760 [.725, .795]	.739 [.710, .767]	.763 [.734, .791]	.736 [.709, .763]	.759 [.733, .786]
Asian	.869 [.769, .969]	.902 [.793, 1.00]	.830 [.711, .950]	.846 [.699, .993]	.847 [.734, .960]	.860 [.721, .999]
Native American	.685 [.618, .753]	.729 [.665, .793]	.706 [.642, .771]	.725 [.662, .787]	.708 [.643, .773]	.732 [.670, .794]
Female violent recidivism						
White	.667 [.590, .743]	.701 [.625, .777]	.680 [.620, .740]	.706 [.649, .764]	.685 [.635, .736]	.707 [.659, .755]
Black	.790 [.690, .891]	.835 [.766, .904]	.783 [.724, .842]	.818 [.765, .872]	.763 [.714, .813]	.810 [.765, .854]
Hispanic	.730 [.639, .822]	.779 [.698, .860]	.728 [.665, .791]	.725 [.660, .791]	.711 [.653, .770]	.718 [.662, .774]
Asian	.521 [.412, .630]	.716 [.618, .813]	.521 [.412, .630]	.716 [.618, .813]	.521 [.412, .630]	.716 [.618, .813]
Native American	.726 [.537, .915]	.743 [.584, .901]	.643 [.507, .778]	.715 [.606, .823]	.590 [.480, .700]	.626 [.520, .732]

Note: White male sample $n = 10,869$; Black male sample $n = 14,750$; Hispanic male sample $n = 7,514$; Asian male sample $n = 582$; Native American male sample $n = 1,421$; white female sample $n = 2,112$; Black female sample $n = 1,237$; Hispanic female sample $n = 1,464$; Asian female sample $n = 96$; Native American female sample $n = 248$.

Table 13 reports the results of the predictive value analyses for the three-year recidivism outcomes by race and ethnic group. The PPVs for the male general recidivism tool indicate that males in the higher RLG recidivated between 66% and 83% of the time; there was variation across racial and ethnic groups, with the higher RLG designation being less accurate for Black and Hispanic individuals at 66% and most accurate for Native American males at 83%. Overall, the violent male PPV was 32%, reflecting the greater difficulty in predicting lower prevalence events (see USDOJ 2021b). With violent outcome PPVs, the tool was most accurate for Black males at 36% and least accurate for Asian males at 24%.

For general recidivism, Native American and Asian females had the highest PPVs at 80% and 91%, and Black and Hispanic females the lowest at 54% and 57%. The female violent PPVs were lower than the other tool PPVs, and more varied across groups, likely a reflection of both the smaller number of overall observations (5,157 females) and the rarity of violent recidivism events by females (i.e., only 285 such events). The female violent tool results accordingly should be interpreted with caution.

For the NPVs, lower risk males did not recidivate between 47% and 86% of the time with an overall mean of 73%; the rate for Native American males at 47% represented the lowest value. This finding suggests that PATTERN is less effective at correctly identifying lower risk Native American males, as this prediction group tended to recidivate at a higher rate compared to the other groups that are classified in the lower RLG. For the violent risk tools, the lower risk designations demonstrated high accuracy, with NPVs ranging from 90% to 96% for males (being lowest for Black and Native American males at 90% each). For females the violent RLGs were accurate 95% of the time overall, ranging from a low of 90% for Native American females to a high of 99% for Asian females.

Table 13. Positive and negative predictive values and false positive and negative rates, by race/ethnicity

<i>Risk assessment tool</i>	<i>PPV</i>	<i>NPV</i>	<i>FPR</i>	<i>FNR</i>
Male general recidivism				
White	0.72	0.74	0.19	0.37
Black	0.66	0.72	0.41	0.23
Hispanic	0.66	0.73	0.23	0.39
Native American	0.83	0.47	0.41	0.25
Asian	0.67	0.86	0.12	0.37
Total	0.68	0.73	0.29	0.30
Male violent recidivism				
White	0.27	0.94	0.29	0.28
Black	0.36	0.90	0.47	0.19
Hispanic	0.28	0.92	0.32	0.31
Native American	0.34	0.90	0.67	0.10
Asian ^a	0.24	0.96	0.20	0.33
Total	0.32	0.92	0.38	0.22
Female general recidivism				
White	0.63	0.73	0.11	0.64
Black	0.54	0.81	0.13	0.57
Hispanic	0.57	0.77	0.09	0.69
Native American ^a	0.80	0.53	0.11	0.64

Asian ^a	0.91	0.93	0.01	0.38
Total	0.61	0.75	0.11	0.64
Female violent recidivism				
White ^a	0.11	0.96	0.05	0.87
Black	0.28	0.96	0.09	0.56
Hispanic ^a	0.12	0.95	0.05	0.87
Native American ^a	0.17	0.90	0.20	0.68
Asian ^a	0.00	0.99	0.05	1.00
Total	0.18	0.95	0.07	0.76

Note: PPV = positive predictive value. NPV = negative predictive value. FPR = false positive rate. FNR = false negative rate. The superscript ^a indicates that at least one of the 2 × 2 cells included fewer than 30 observations, so the generalizability to population estimates is less certain due to small sample size. Full distribution tables are provided in Appendix D. White male sample $n = 10,869$; Black male sample $n = 14,750$; Hispanic male sample $n = 7,514$; Native American male sample $n = 1,421$; Asian male sample $n = 582$; white female sample $n = 2,112$; Black female sample $n = 1,237$; Hispanic female sample $n = 1,464$; Native American female sample $n = 248$; Asian female sample $n = 96$.

Table 14 reports the differential prediction analyses for the 2019 cohort. The results (Table 14 and Figure 4) show that the Black male relative overprediction documented in previous reports (USDOJ 2021b; 2023) continued to worsen, with Black males now over 6% less likely to engage in general recidivism compared to white males holding constant the PATTERN score, while the relative Hispanic male overprediction mitigated slightly from about 6% (USDOJ 2023) to just over 4%. As in previous years, Native American risk was significantly underpredicted compared to white males on the general recidivism tool (see USDOJ 2021b; 2023). The violent male risk tool exhibited results that were mostly not statistically significant, with one exception being a nearly 5% underprediction of violent recidivism for Black males compared to white males (i.e., counter to the findings on the general tool, when controlling for the PATTERN score, Black males were more likely to have a violent recidivism event compared to white males). For the violent male tool, the Hispanic results were statistically significant with a modest underprediction of 1.3%.

The overprediction of Black, Hispanic, and Asian females relative to white females on the general recidivism score was similar to the previous validation and revalidation samples (USDOJ 2021b; 2023) at 7.9%, 4.4%, and 9.9%, respectively, though the Black female impact continues to worsen (from 6.9% among the FY 2018 cohort to 7.9% for this FY 2019 cohort). Native American female general risk was underpredicted by about 11.1% compared to white females, also consistent with previous years. Finally, female violent results were mostly not statistically significant and similar to previous years, though the Black violent underprediction of 1.6% was statistically significant this year. The Model 4 interaction effects were still mostly not statistically significant, with a few exceptions for Asian general male and Black violent female.²⁴

²⁴ The differential prediction results were also conducted for the one- and two-year follow-up outcomes (not shown), and the findings were substantively similar to those found with the three-year follow-up.

Table 14. Summary of differential prediction regression analyses findings

	Model 3				Model 4			
	<i>AME</i>	<i>OR</i>	<i>p</i>		<i>AME</i>	<i>OR</i>	<i>p</i>	
General Male								
Black	-6.63%	0.71	0.000	***	-6.51%	0.66	0.000	***
Hispanic	-4.30%	0.80	0.000	***	-4.07%	0.71	0.000	***
Native American	12.53%	1.96	0.000	***	12.97%	2.03	0.000	***
Asian	-9.31%	0.61	0.000	***	-6.48%	0.32	0.000	***
Black x score						1.00	0.229	
Hispanic x score						1.00	0.085	
Native American x score						1.00	0.873	
Asian x score						1.02	0.002	**
Violent Male								
Black	4.93%	1.46	0.000	***	4.94%	1.46	0.000	***
Hispanic	1.31%	1.11	0.019	*	1.38%	1.05	0.670	
Native American	1.36%	1.12	0.115		2.02%	1.50	0.094	
Asian	-2.09%	0.83	0.241		-1.40%	0.58	0.155	
Black x score						1.00	0.955	
Hispanic x score						1.00	0.587	
Native American x score						0.99	0.221	
Asian x score						1.01	0.291	
General Female								
Black	-7.86%	0.63	0.000	***	-7.83%	0.69	0.020	*
Hispanic	-4.39%	0.78	0.002	**	-4.39%	0.78	0.108	
Native American	11.13%	1.81	0.000	***	10.76%	1.68	0.113	
Asian	-9.85%	0.55	0.058		-7.86%	0.30	0.044	*
Black x score						1.00	0.500	
Hispanic x score						1.00	0.928	
Native American x score						1.00	0.805	
Asian x score						1.03	0.187	
Violent Female								
Black	1.63%	1.39	0.039	*	1.47%	0.94	0.814	
Hispanic	0.85%	1.20	0.257		0.90%	1.14	0.564	
Native American	2.20%	1.53	0.075		4.09%	2.65	0.010	*
Asian	-3.33%	0.28	0.211		-3.49%	0.35	0.319	
Black x score						1.05	0.046	*
Hispanic x score						1.01	0.722	
Native American x score						0.95	0.160	
Asian x score						0.94	0.672	

Note: AME = average marginal effects. OR = odds ratio. White individuals are the referent group. *** $p \leq 0.001$. ** $p \leq 0.01$. * $p \leq 0.05$. White male sample $n = 10,869$; Black male sample $n = 14,750$; Hispanic male sample $n = 7,514$; Native American male sample $n = 1,421$; Asian male sample $n = 582$; white female sample $n = 2,112$; Black female sample $n = 1,237$; Hispanic female sample $n = 1,464$; Native American female sample $n = 248$; Asian female sample $n = 96$.

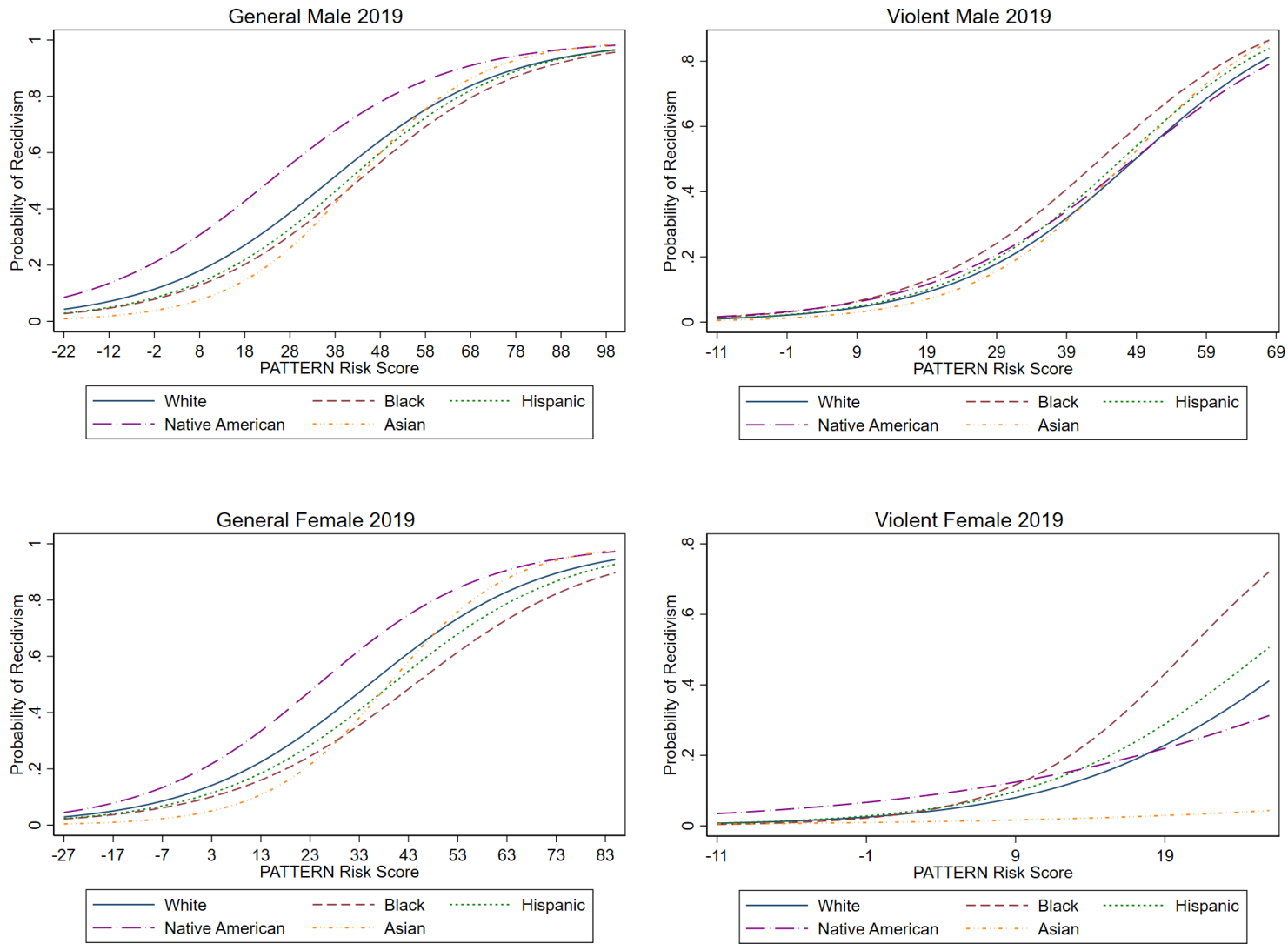


Figure 4. Differential prediction plots, by PATTERN instrument

Part 6: Discussion and Conclusion

This study represents the fourth review and revalidation report on PATTERN. As mandated by the FSA, the current study evaluated PATTERN for its predictive accuracy, dynamic validity, and racial/ethnic neutrality on a subsequent cohort of FY 2019 FBOP releasees. The current study findings continue to demonstrate that PATTERN is a strong predictor of general and violent recidivism at one-, two, and three-year follow-up periods, with AUCs ranging from .745 to .776. Comparisons of recidivism rates by RLC and predictive value analyses by RLG also continue to indicate that such risk level designations provide meaningful distinctions of recidivism risk. In addition, the results continue to suggest that individuals can change their risk scores and levels during confinement. Such changes in risk were not exclusively driven by changes in age. Those who reduced their RLC from first to last assessment were generally shown to have the lowest recidivism rates, followed by those who maintained the same risk level and those with a higher risk level, respectively. While the study findings indicate that PATTERN is predictively accurate across the five racial and ethnic groups analyzed, there remains evidence that the instruments overpredict the risk of recidivism for some racial and ethnic groups relative to white individuals (e.g., Black, Hispanic, and Asian males and females on the general tools), as was true of previous versions of PATTERN and has been documented in prior reports (USDOJ 2021a, 2021b, 2023).

Consistent with the previous review and revalidation reports, these findings also continue to document meaningful differences across the RLC-by-race distributions. Differences in group risk level distributions can be referred to as “differential impact.” These distributions do not consider the accuracy or parity of recidivism predictions, only group differences in risk categories. Differential impact is distinguished from differential prediction, in which members of different groups with the same risk scores have different rates of recidivism. Differential prediction (discussed further below) indicates that a tool predicts differently for different racial and ethnic groups, a form of racial bias.

Differential impact and differential prediction can both be affected by biased data, and thus might be mitigated, for example, through the selection of prediction or outcome items, where alternative data sources are available. As with all administrative data sources, it is possible that there are biases in PATTERN data sources and its measures, and NIJ will continue to work to explore and mitigate these biases with its partners. Even if these sources of data bias could be identified and corrected, however, there may still be some group-based differences that are not attributable to data bias. If so, groups may experience different risk scores and categories that would not necessarily indicate bias. Further, it is often difficult (or impossible) to discern whether some observed group-level differences in data are genuine or reflect some sort of systemic bias. It would be possible for some group-level differences to exist, and for a tool to achieve parity in prediction across groups, while still demonstrating differential impact.

NIJ recognizes concerns about potential bias in risk factors themselves (such as criminal history) and in the recidivism outcome measure, which captures arrest rather than conviction. NIJ is committed to the ongoing efforts to explore additional data points, including the possibility of a conviction-based recidivism outcome measure. As noted above, FBOP is actively pursuing a way to capture reconviction information more accurately and comprehensively through a new partnership with the Data Science Discovery Program at the University of California, Berkeley.

Although racial and ethnic neutrality can be examined through numerous metrics,²⁵ the racial and ethnic fairness analyses presented here have prioritized the differential prediction findings, reflecting the current emphasis in the field (e.g., Skeem & Lowenkamp 2016). While an effective tool might still fairly reflect group-based differences in risk categorization, an unbiased tool should predict similarly across racial and ethnic groups. To address these questions, the study examined AUCs and predictive values by race, and employed regression analyses to test for differential prediction. The findings indicate that different risk scores correspond to different recidivism likelihoods across racial and ethnic groups — evidence of differential prediction.²⁶

This disparity remains NIJ’s leading concern related to PATTERN, and one which it is committed to addressing. Overall, the differential prediction results are consistent with previous years and thus mirror the concerns raised in the USDOJ (2021b and 2023) reports. Those reports discussed the inherent impossibility of satisfying all notions of racial and ethnic fairness, since different definitions are interrelated and conflicting (USDOJ 2021b, 44-45; see also Berk et al. 2021, Chouldechova 2017). In addition, while the focus here is on differential results adapted from the *Standards for Educational and Psychological Testing*, those standards do not impose strict requirements on absolute parity across groups (see USDOJ 2021b, 45). Furthermore, as discussed in USDOJ (2021b, 45), PATTERN addresses five distinct racial and ethnic groups, which poses unique challenges over the examinations found in the criminal justice literature that have typically considered just two racial or ethnic groups. Nevertheless, the differential prediction results raise a clear concern related to PATTERN’s racial and ethnic neutrality. As noted in USDOJ (2021b), “there are no simple solutions to this complex problem” and indicated that “deliberate study and engagement with stakeholders and experts are warranted to identify an optimal path forward” (46). NIJ and its consultants continue to investigate potential solutions for the differential prediction analyses identified. In 2023, Dr. Greg Ridgeway (University of Pennsylvania) joined the review and revalidation team as a statistical consultant; NIJ continues to explore methodological ways to fulfill the FSA’s mandate “to ensure that any disparities identified ... are reduced to the greatest extent possible” (FSA § 3631 (b)(5)).

Finally, the current study cohort included individuals released through September 30, 2019. Next year’s review and revalidation report will cover October 1, 2019, through September 30, 2020, which is significant for two reasons. First, to date, the review and revalidation analyses have been retrospective, conducted on pre-deployment cohorts as individuals were released from custody and into the community with a three-year recidivism follow-up time. PATTERN, and the FSA additional earned time rules, went into place in January 2020, meaning the upcoming cohort will be the first for which individuals were scored on PATTERN and may have earned additional time credit prior to release based on their scores. Second, early 2020 also marked the onset of the COVID-19 pandemic which may have affected early release decisions apart from the new FSA rules. In addition, the combination of COVID-19 precautions and the civil unrest that marked the summer of 2020 may have had an impact on policing and arrest patterns, which

²⁵ As noted in USDOJ (2021b), when base rates of recidivism differ, it is impossible to achieve parity in PPVs/NPVs and FNRs/FPRs (Berk et al. 2021; Chouldechova 2017). Goel et al. (2021, 16) note that “differences in false positive rates often tell us more about the underlying populations than about bias in the algorithm.”

²⁶ These results may be influenced by changes in arrest rates by groups over time. However, a time trend does not fully account for the Black female differential prediction, as statistically significant overprediction existed in the development training and validation samples (USDOJ 2021b).

could directly or indirectly affect the recidivism outcomes for the revalidation analyses. Thus, the upcoming review and revalidation would be of particular interest as the first partially-prospective PATTERN cohort, but will also be tempered with caution due to the potential confounding effects of the follow-up period.

References

- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods and Research*, 50(1), 3-44.
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153-163.
- Federal Bureau of Prisons (FBOP). (2006). *Inmate security designation and custody classification* (policy number 5100.08). U.S. Department of Justice.
- Federal Bureau of Prisons (FBOP). (2019). *Inmate security designation and custody classification* (policy number 5100.08, CN-1). U.S. Department of Justice.
- Federal Bureau of Prisons (FBOP). (2021). *Evidence-based recidivism reduction (EBRR) programs and productive activities (PA)*. U.S. Department of Justice.
- Federal Bureau of Prisons (FBOP). (2022a, March). *First Step Act: Initial review of the SPARC-13 needs assessment system*. U.S. Department of Justice.
- Federal Bureau of Prisons (FBOP). (2022b, January). *FSA time credits*. U.S. Department of Justice.
- Goel, S., Shroff, R., Skeem, J., & Slobogin, C. (2021). The accuracy, equity, and jurisprudence of criminal risk assessment. In R. Vogl (Ed.), *Research Handbook on Big Data Law*. Elgar.
- Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. *Criminology*, 54(4), 680-712.
- U.S. Department of Justice (USDOJ). (2019, July). *The First Step Act of 2018: Risk and needs assessment system*. Office of the Attorney General.
- U.S. Department of Justice (USDOJ). (2020, January). *The First Step Act of 2018: Risk and needs assessment system – Update*. Office of the Attorney General.
- U.S. Department of Justice (USDOJ). (2021a, January). *2020 review and revalidation of the First Step Act risk assessment tool*. National Institute of Justice.
- U.S. Department of Justice (USDOJ). (2021b, December). *2021 review and revalidation of the First Step Act risk assessment tool*. National Institute of Justice.
- U.S. Department of Justice (USDOJ). (2022a, April). *First Step Act annual report*. Office of the Attorney General.
- U.S. Department of Justice (USDOJ). (2023, March). *2022 review and revalidation of the First Step Act risk assessment tool*. National Institute of Justice.

Appendix A

Description of PATTERN's static and dynamic risk items

Current age	Years between assessment and date of birth, rounded down; converted into six ordinal categories: <i>25 and younger, 26 to 29, 30 to 40, 41 to 50, 51 to 60, or 61 and older.</i>
Walsh with conviction	Identification as a sex offender based on the Adam Walsh Act criteria (see 34 U.S.C. § 20911).
Violent offense	A current conviction for a violent offense, including but not limited to firearms violations, homicide, child abuse, robbery, sex trafficking, and sexual assault (see FBOP 2020).
Criminal history points	The number of criminal history points taken (from the most recent BRAVO); converted into six ordinal categories that match those used by the United States Sentencing Commission: <i>0 to 1 point, 2 to 3 points, 4 to 6 points, 7 to 9 points, 10 to 12 points, or 13 or more points.</i>
History of escapes	Years since last escape attempt by seriousness (from the most recent BRAVO); converted into four ordinal categories: <i>None, >10 years minor, 5 to 10 years minor, or less than 5 years minor or any serious.</i>
History of violence	Number of years from last act of violence by seriousness (from the most recent BRAVO); converted into eight ordinal categories: <i>None, greater than 10 years minor, greater than 15 years serious, 5 to 10 years minor, 10 to 15 years serious, less than 5 years minor, 5 to 10 years serious, or less than 5 years serious.</i>
Education status	The highest grade level completed; converted into three ordinal categories: <i>High school degree or GED, enrolled and progressing in GED program, or no verified degree and not participating in GED program.</i>
Drug program status	This measure combines two sources of information: (1) Identification of substance abuse history (from the most recent BRAVO) and (2) completion of residential or non-residential drug programming during the current incarceration. ^a This variable is then converted into four ordinal categories: <i>No drug need indicated, completed residential drug treatment, completed non-residential drug treatment, or need indicated but no drug treatment completed.</i>
All incident reports	Number of guilty incident reports within the past 120 months following one's incarceration date; includes only incident reports, not individual acts if multiple acts were committed in a single incident; excludes incident reports occurring during pretrial, holdover, or from prior FBOP incarcerations; converted into four ordinal categories: <i>No incident, 1 incident, 2 incidents, or 3 or more incidents.</i>
Serious incident reports	Number of guilty serious and violent incident reports within the past 120 months following one's incarceration date; includes only incident reports, not individual acts if multiple acts were committed in a single incident; excludes incident reports occurring during pretrial, holdover, or from prior FBOP incarcerations; converted into four ordinal categories: <i>No incident, 1 incident, 2 incidents, or 3 or more incidents.</i>
Time since last incident report	Number of months between the assessment date and the date of the most recent incident report, rounded down; converted into four ordinal categories: <i>12+ months or no incident, 7 to 12 months, 3 to 6 months, or less than 3 months.</i>
Time since last serious incident report	Number of months between the assessment date and the date of the most recent serious or violent incident report, rounded down; converted into four ordinal categories: <i>12+ months or no incident, 7 to 12 months, 3 to 6 months, or less than 3 months.</i>
Financial responsibility refuse	Noncompliance with financial responsibility during incarceration for payment toward victim restitution and dependents.
Programs completed	Number of ACE, BRAVE, Challenge, Drug Education, Life Connections, Parenting, Skills, Sex Offender Residential Treatment, Sex Offender Non-Residential Treatment, STAGES, and Step Down courses successfully completed during the current incarceration; converted into five ordinal categories: <i>No program, 1 program, 2 to 3 programs, 4 to 10 programs, or 11 or more programs.</i> ^b
Work programs completed	Number of technical and vocational courses completed during the current incarceration; federal industry employment (UNICOR) is counted as a program completion if the individual worked

	at least one day; converted into three ordinal categories: <i>No program, 1 program, or 2 or more programs.</i>
--	---

Notes: ^a This measure does not include all the evidence-based recidivism reduction (EBRR) drug programs and other drug-related productive activities (PAs) currently available throughout FBOP, as these data were not available during the study observation period (see FBOP 2021).

^b This measure does not include all the EBRR programs and PAs currently available in FBOP, as these data were not available during the study observation period. Additionally, some of the programs currently included in this variable, such as ACE, are not considered EBRR programs or PAs by FBOP policy, though they predictively correlate with recidivism (see FBOP 2021).

Appendix B

PATTERN Scoring Guide

Item	Category	General Male	Violent Male	General Female	Violent Female
1. Current age	<i>> 60</i>	0	0	0	0
	<i>51-60</i>	7	4	6	1
	<i>41-50</i>	14	8	12	2
	<i>30-40</i>	21	12	18	3
	<i>26-29</i>	28	16	24	4
	<i>< 26</i>	35	20	30	5
2. Walsh with conviction	<i>No</i>	0			
	<i>Yes</i>	2			
3. Violent offense	<i>No</i>	0	0	0	0
	<i>Yes</i>	5	7	1	3
4. Criminal history points	<i>0-1</i>	0	0	0	0
	<i>2-3</i>	8	3	8	1
	<i>4-6</i>	16	6	16	2
	<i>7-9</i>	24	9	24	3
	<i>10-12</i>	32	12	32	4
	<i>13+</i>	40	15	40	5
5. History of escapes	<i>None</i>	0	0	0	0
	<i>> 10 years minor</i>	3	2	3	1
	<i>5-10 years minor</i>	6	4	6	2
	<i>< 5 years minor or any serious</i>	9	6	9	3
6. History of violence	<i>None</i>	0	0	0	0
	<i>> 10 years minor</i>	1	2	1	1
	<i>> 15 years serious</i>	2	4	2	2
	<i>5-10 years minor</i>	3	6	3	3
	<i>10-15 years serious</i>	4	8	4	4
	<i>< 5 years minor</i>	5	10	5	5
	<i>5-10 years serious</i>	6	12	6	6
	<i>< 5 years serious</i>	7	14	7	7
7. Education status	<i>Not enrolled</i>	0	0	0	0
	<i>Enrolled in GED</i>	-1	-1	-3	-1
	<i>High school degree/GED</i>	-2	-2	-6	-2
8. Drug program status	<i>Need indicated/no completion</i>	0	0	0	0
	<i>Completed non-residential drug treatment</i>	-2	-1	-3	-1
	<i>Completed residential drug treatment</i>	-4	-2	-6	-2
	<i>No need indicated</i>	-6	-3	-9	-3
9. All incident reports	<i>0</i>	0	0	0	
	<i>1</i>	1	1	1	
	<i>2</i>	2	2	2	
	<i>3+</i>	3	3	3	
10. Serious incident reports	<i>0</i>	0	0		
	<i>1</i>	1	1		
	<i>2</i>	2	2		
	<i>3+</i>	3	3		
11. Time since last incident report	<i>12+ months</i>	0	0	0	0
	<i>7-12 months</i>	1	1	2	1
	<i>3-6 months</i>	2	2	4	2
	<i>< 3 months</i>	3	3	6	3

12. Time since last serious incident report	<i>12+ months</i>			0	0
	<i>7-12 months</i>			1	1
	<i>3-6 months</i>			2	2
	<i>< 3 months</i>			3	3
13. Financial responsibility refuse	<i>No</i>	0		0	0
	<i>Yes</i>	2		3	1
14. Programs completed	<i>0</i>	0	0	0	0
	<i>1</i>	-3	-1	-2	-1
	<i>2-3</i>	-6	-2	-4	-2
	<i>4-10</i>	-9	-3	-6	-3
	<i>11+</i>	-12	-4	-8	-4
15. Work programs completed	<i>0</i>	0	0	0	0
	<i>1</i>	-1	-1	-2	-1
	<i>2+</i>	-2	-2	-4	-2
Risk level category	<i>Minimum</i>	-22 to 5	-11 to 7	-27 to 7	-11 to 1
	<i>Low</i>	6 to 39	8 to 24	8 to 38	2 to 11
	<i>Medium</i>	40 to 54	25 to 31	39 to 52	12 to 17
	<i>High</i>	55 to 109	32 to 71	53 to 102	18 to 30

Appendix C

PATTERN areas under the curve and 95% confidence intervals, by assessment type and follow-up period (Earned time credit eligible sample only)

Risk assessment Recidivism follow-up period	AUC [95% CI] First Assessment	AUC [95% CI] Last Assessment
Male general recidivism		
One-year follow-up	.705 [.696, .713]	.753 [.746, .761]
Two-year follow-up	.722 [.715, .729]	.768 [.761, .775]
Three-year follow-up	.727 [.720, .734]	.769 [.763, .776]
Male violent recidivism		
One-year follow-up	.718 [.703, .732]	.749 [.735, .763]
Two-year follow-up	.725 [.714, .736]	.752 [.741, .762]
Three-year follow-up	.723 [.714, .733]	.752 [.743, .761]
Female general recidivism		
One-year follow-up	.732 [.714, .750]	.771 [.754, .788]
Two-year follow-up	.742 [.726, .757]	.775 [.760, .790]
Three-year follow-up	.753 [.738, .767]	.780 [.766, .794]
Female violent recidivism		
One-year follow-up	.699 [.639, .758]	.755 [.702, .808]
Two-year follow-up	.705 [.664, .745]	.737 [.699, .776]
Three-year follow-up	.701 [.668, .734]	.730 [.698, .762]

Note: Male sample $n = 20,106$; female sample $n = 4,521$.

Appendix D

Risk level categories and recidivism by race and ethnicity, FY 2019 revalidation sample, one-year follow-up period

	Male General		Male Violent		Female General		Female Violent	
	<i>Did Not Recidivate</i>	<i>Recidivated</i>	<i>Did Not Recidivate</i>	<i>Recidivated</i>	<i>Did Not Recidivate</i>	<i>Recidivated</i>	<i>Did Not Recidivate</i>	<i>Recidivated</i>
White								
Minimum	2006	149	3051	18	584	19	992	8
Low	3797	787	3930	119	868	226	974	25
Medium	1254	833	1329	116	190	96	97	6
High	859	1184	1984	332	65	64	10	0
Black								
Minimum	568	14	998	6	376	9	584	2
Low	4706	753	5261	244	516	77	505	10
Medium	2967	1306	2678	262	110	43	102	8
High	2201	2235	4267	1034	57	49	23	3
Hispanic								
Minimum	646	23	1111	6	372	16	715	4
Low	3298	599	3451	105	736	119	642	21
Medium	1099	595	1143	70	103	47	69	4
High	566	688	1374	254	37	34	7	2
Native American								
Minimum	32	2	50	0	22	3	48	0
Low	302	157	314	14	113	49	143	4
Medium	190	263	213	10	20	18	38	4
High	153	322	678	142	9	14	11	0
Asian								
Minimum	144	2	215	1	53	1	65	0
Low	253	32	217	7	27	4	25	1
Medium	59	16	57	2	1	6	5	0
High	36	40	72	11	1	3	0	0

Risk level categories and recidivism by race and ethnicity, FY 2019 revalidation sample, two-year follow-up period

	Male General		Male Violent		Female General		Female Violent	
	<i>Did Not Recidivate</i>	<i>Recidivated</i>	<i>Did Not Recidivate</i>	<i>Recidivated</i>	<i>Did Not Recidivate</i>	<i>Recidivated</i>	<i>Did Not Recidivate</i>	<i>Recidivated</i>
White								
Minimum	1955	200	3036	33	571	32	986	14
Low	3383	1201	3807	242	763	331	958	41
Medium	942	1145	1247	198	154	132	92	11
High	546	1497	1736	570	45	84	10	0
Black								
Minimum	559	23	991	13	369	16	580	6
Low	4205	1254	5034	471	472	121	491	24
Medium	2317	1956	2433	507	92	61	92	18
High	1455	2981	3557	1744	43	63	18	8
Hispanic								
Minimum	636	33	1105	12	365	23	707	12
Low	2970	927	3330	226	651	204	624	39
Medium	855	839	1055	158	83	67	65	8
High	387	867	1226	402	26	45	7	2
Native American								
Minimum	32	2	50	0	22	3	47	1
Low	244	215	303	25	89	73	138	9
Medium	128	325	203	20	12	26	35	7
High	89	386	579	241	5	18	10	1
Asian								
Minimum	143	3	215	1	52	2	65	0
Low	239	46	212	12	27	4	25	1
Medium	44	31	54	5	1	6	5	0
High	20	56	58	25	1	3	0	0

Risk level categories and recidivism by race and ethnicity, FY 2019 revalidation sample, three-year follow-up period

	Male General		Male Violent		Female General		Female Violent	
	<i>Did Not Recidivate</i>	<i>Recidivated</i>	<i>Did Not Recidivate</i>	<i>Recidivated</i>	<i>Did Not Recidivate</i>	<i>Recidivated</i>	<i>Did Not Recidivate</i>	<i>Recidivated</i>
White								
Minimum	1901	254	3025	44	558	45	982	18
Low	3102	1482	3698	351	676	418	937	62
Medium	771	1316	1169	276	118	168	91	12
High	401	1642	1565	741	37	92	10	0
Black								
Minimum	543	39	986	18	361	24	577	9
Low	3836	1623	4841	664	428	165	475	40
Medium	1911	2362	2245	695	82	71	81	29
High	1083	3353	3020	2281	37	69	17	9
Hispanic								
Minimum	624	45	1100	17	354	34	704	15
Low	2698	1199	3220	336	603	252	611	52
Medium	712	982	989	224	71	79	65	8
High	305	949	1062	566	23	48	7	2
Native American								
Minimum	30	4	49	1	22	3	45	3
Low	202	257	290	38	78	84	131	16
Medium	100	353	184	39	9	29	34	8
High	60	415	507	313	3	20	10	1
Asian								
Minimum	142	4	213	3	52	2	65	0
Low	229	56	210	14	27	4	25	1
Medium	35	40	52	7	0	7	5	0
High	15	61	56	27	1	3	0	0