U.S. Department of Justice
Office of Justice Programs
National Institute of Justice

National Institute of Justice

2022 Review and Revalidation of the First Step Act Risk Assessment Tool

March 2023

NCJ 305720

NIJ | National Institute of Justice

STRENGTHEN SCIENCE. ADVANCE JUSTICE.

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
Strengthen Science • Advance Justice
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 advance scientific research, development, and evaluation to enhance the administration of justice and public safety.

The National Institute of Justice is a component 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.

Table of Contents

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 (BOP). The FSA also required that the U.S. Department of Justice (DOJ) review, validate, and publicly release the risk and needs assessment system — the Prisoner Assessment Tool Targeting Estimated Risk and Needs (PATTERN) — 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 PATTERN. This document is the third review and revalidation report, following USDOJ (2021a) and USDOJ (2021b). The current report analyzes a subsequent cohort of fiscal year (FY) 2018 BOP releasees and evaluates PATTERN for its predictive accuracy, dynamic validity, and racial and ethnic neutrality, as mandated by the FSA. It also expands upon the prior analyses by including one-, two-, and three-year recidivism outcomes, assessing what proportion of change in risk scores and levels is influenced by the current age item, and providing additional descriptive information on individual items, risk scores and levels, and outcomes by race and ethnic group. Finally, this report provides an update on actions taken by NIJ and DOJ in the past year and the ongoing efforts to review and improve PATTERN.

The FY 2018 cohort study findings continue to demonstrate that PATTERN is a strong and valid predictor of general and violent recidivism, with Area Under the Curve (AUC) statistics ranging from 0.76 to 0.78. PATTERN predicted both types of recidivism at the one-, two-, and three-year follow-up periods. 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, the results continue to suggest that individuals can change their risk scores and levels during confinement. Changes in risk were not driven exclusively by changes in age. 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 study findings continue to indicate that PATTERN is predictively accurate across the five racial and ethnic groups analyzed, there remains evidence that the instrument overpredicts the risk of recidivism for some racial and ethnic groups relative to white individuals (e.g., Black, Hispanic, and Asian individuals on the male and female general tools).

NIJ and DOJ remain committed to revising and updating PATTERN to improve the "equitability, efficiency, and predictive validity of the risk assessment system" (USDOJ 2021b, 47), including "to ensure that racial disparities are reduced to the greatest extent possible, as required by the FSA, 18 U.S.C. § 3631(b)(5)" (USDOJ 2022, 11). In April 2022, the U.S. Attorney General modified the general recidivism RLC cutpoints to reduce racial and ethnic disparities in PATTERN and expand the number of individuals eligible for FSA benefits. Throughout 2022, NIJ has also engaged with subject matter experts to discuss potential strategies for addressing racial and ethnic bias, and DOJ held stakeholder engagement sessions to solicit additional feedback on PATTERN more generally. These engagements have informed the ongoing efforts to refine and improve PATTERN, including considerations of how recidivism is defined, which data inputs are used for scoring, and whether modeling changes might reduce the racial and ethnic disparities in the tool. NIJ and its consultants are dedicated to continuing this work with experts and stakeholders to address these important issues.

Part 1: Background and 2022 Developments

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 (BOP). The FSA also required the U.S. Department of Justice (DOJ) to review, validate, and publicly release the risk and needs assessment system — the Prisoner Assessment Tool Targeting Estimated Risk and Needs (PATTERN) — 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 PATTERN. This document is the third review and revalidation report, following USDOJ (2021a) and USDOJ (2021b).[1] The USDOJ (2021a) report documented discrepancies in the scoring, coding, and specification of the PATTERN risk items which required reconstruction of the tool.[2]

The new version, PATTERN version 1.3, was formally adopted in April 2022 (USDOJ 2022). The USDOJ (2021b) report published revalidation results on the predictive accuracy, dynamic validity, and racial and ethnic neutrality of PATTERN version 1.3 as mandated by the FSA. Those results showed 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; i.e., minimum-, low-, medium-, and high-risk). Although the study found that PATTERN was a strong predictor of recidivism across the five racial and ethnic groups analyzed, the results also included 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]

---

[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, BOP 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 BOP 2022).

[2] After an independent review of the data, syntax files, and other material used to develop PATTERN, the first review and revalidation report identifying several discrepancies in the scoring, coding, and specification of the risk assessment items was released by NIJ in January 2021. Following this discovery, BOP 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] 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 BOP'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).

Thus, transitioning to PATTERN version 1.3 neither exacerbated nor solved these racial bias issues.

The USDOJ (2021b) report indicated that NIJ and its consultants would collaborate with DOJ subject matter experts, BOP, and stakeholders to "explore whether further refinements to items and the scoring scheme of PATTERN may help improve the equitability, efficiency, and predictive validity of the risk assessment system" (47). USDOJ (2021b) acknowledged that "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). The report also stated that NIJ and its consultants would "continue to investigate potential solutions for the differential prediction issues identified" (USDOJ 2021b, 47). Efforts to address these issues to date can be grouped into three primary initiatives:

(1) **New Risk Level Category (RLC) Cutpoints.** The most significant impact of PATTERN lies in an individual's assignment to a risk level grouping (RLG). Those in the lower RLG (i.e., minimum and low RLCs) are statutorily eligible to earn up to five additional days of earned time credit toward early release per month, while those in the higher RLG (i.e., medium and high RLCs) are not.[4] NIJ and its consultants conducted analyses to inform the Attorney General's adjustment of the RLC cutpoints for the male and female general tools. The RLC cutpoints were adjusted "(1) to enhance opportunities for eligible individuals to earn time credits … (2) to help mitigate the effects of various racial and ethnic disparities … and (3) to continue to ensure that the PATTERN's risk level designations promote public safety and the reduction of recidivism" (USDOJ 2022, 14). More specifically, the thresholds separating the low- and medium-risk general recidivism categories were raised in both the male and female models to make more individuals eligible for additional earned time credit. In addition to making more people eligible generally, this strategy made a greater proportion of minority individuals eligible, thereby reducing racial disparities. As USDOJ (2022, 15) documents, the racial and ethnic disparities were improved for all groups for both the male and female general tools because of these changes (except for the white to Hispanic ratio under the female tool, which remained unchanged). The report acknowledges that this change "will not directly correct the racial differential prediction rates found to be associated with the PATTERN tool, including version 1.3. But making these cut point adjustments for the general tool is an important step towards mitigating the racial disparities associated with the tool" (USDOJ 2022, 17).

(2) **Review of PATTERN's Data and Methodology**. NIJ and its consultants continue to investigate new modeling strategies to mitigate differential prediction in the risk scores. Following the April 2022 modifications to PATTERN, NIJ engaged with subject matter experts across several disciplines to develop potential strategies for mitigating racial and ethnic disparities. These experts included: Dr. Ashlee Barnes-Lee (research associate,

[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).

School of Social Work, Michigan State University); Dr. Sharad Goel (professor of public policy, Harvard Kennedy School); Aziz Huq (professor, University of Chicago Law School); Dr. Christopher Lowenkamp (social science analyst, Administrative Office of the U.S. Courts); Sandra Mayson (professor, University of Pennsylvania Carey Law School); Dr. Greg Ridgeway (chair of criminology and professor of statistics and data science, University of Pennsylvania); Dr. Jennifer Skeem (professor of public policy and social welfare, University of California, Berkeley); and Dr. Christopher Sullivan (professor and chair of the Department of Criminology and Criminal Justice, University of Missouri — St. Louis). The feedback from these experts covered a variety of topics, including the long-term need for additional data collection by BOP, strategies for refining the current approach through different operationalizations of measures and a revised training and validation sampling methodology, and further exploration of alternative statistical modeling techniques. NIJ and its consultants are currently pursuing remodeling strategies informed by these discussions.

(3) **Stakeholder Engagements**. DOJ solicited additional feedback on PATTERN through two stakeholder engagement sessions held in September 2022. The stakeholders included: Brandon Buskey (director of the Criminal Law Reform Project, American Civil Liberties Union), Patti Butterfield (adjunct faculty, Southern New Hampshire University), Cortland Broyles (federal affairs manager, Justice Action Network), Matthew DeMichele (director of the Center for Courts and Corrections Research, RTI International), Liz Komar (sentencing reform counsel, The Sentencing Project), Faye Taxman (professor and director of the Center for Advancing Correctional Excellence, George Mason University), Jim Felman (chair of the First Step Act Implementation Task Force, Criminal Justice Section, American Bar Association), Ames Grawert (senior counsel, Brennan Center for Justice), Grant Duwe (director of research and evaluation, Minnesota Department of Corrections), Elizabeth Blackwood (counsel and project director, National Association of Criminal Defense Lawyers), Patricia Richman (counsel, National Sentencing Resource Counsel, Federal Public and Community Defenders), Shawn Bushway (senior policy researcher, RAND Corporation), Lauren Brinkley-Rubenstein (associate professor of social medicine, Duke University), James Byrne (professor emeritus, University of Massachusetts Lowell), and Melissa Hamilton (professor of law and criminal justice, University of Surrey School of Law). The written and verbal feedback from these stakeholders urged DOJ to consider several recommendations, including considering a narrower definition of recidivism (e.g., one based on conviction outcomes rather than arrests), increasing transparency, adding additional dynamic factors, exploring a decay factor for criminal history, and reconsidering the inclusion of multiple disciplinary infraction measures. As NIJ and its consultants continue to pursue mitigation of racial and ethnic disparities through remodeling strategies, we will also consider the feasibility of additional refinements to PATTERN based on the stakeholder engagement feedback.

In addition to these targeted efforts at improving PATTERN, the FSA mandates the annual review and revalidation of the risk and needs assessment system. Accordingly, the current report documents the revalidation of PATTERN on a subsequent cohort of incarcerated persons — those released from BOP custody during fiscal year (FY) 2018. 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 2021b*)*. It also expands that investigation in three important ways. First, this report conducts additional predictive analyses with one-, two-, and three-year post-release recidivism outcomes. Second, this report expands its dynamic analyses by assessing to what extent changes in risk scores and levels are influenced by updates to the current age item. Finally, this report provides additional descriptive information on individual items, risk scores and levels, and outcomes by race and ethnic group.

Part 2: Method

Data and Sample

The sampling frame for this study was all individuals released from BOP custody between October 1, 2017, and September 30, 2018 (FY 2018) who did not have an active detainer in place upon release and were not known to have died within three years of release ($N = 39{,}991$). Individuals were excluded from the study if their initial BOP classification assessment — the Bureau Risk and Verification Observation (BRAVO) — was not completed under the current version and policy (i.e., 5100.08; see BOP 2006). This criterion was necessary because several of the current BRAVO items are needed to score PATTERN. The criterion mostly excluded individuals who were admitted to BOP custody prior to September 2006.[5] There were 37,874 individuals who met the study inclusion criteria, including 32,804 males and 5,070 females.

The dataset analyzed in this study was obtained from the BOP's Office of Research and Evaluation. 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.[6] 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.[7] 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.[8] Those four RLCs were mandated by the FSA. In practice, BOP 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.[9]

There are two types of post-release recidivism analyzed in this study. General recidivism is defined as any return to BOP custody or rearrest within three years of release from BOP 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 act of

[5] There were 2,117 individuals (or 5.3% of the sample) who were excluded due to this criterion.
[6] 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 BOP 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.
[7] For more information on the construction of PATTERN version 1.3, see USDOJ (2021b).
[8] Note that the Attorney General modified the male and female general risk cutpoints in April 2022 (see USDOJ 2022).
[9] For more information about the earned time credits rule, see BOP (2022b).

violence[10] within three years of release from BOP 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. 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, followed by 41-50, 26-29, 51-60, < 26, and > 60. Approximately 11% of the total sample had a Walsh (sex offender) conviction and about a third were convicted of a violent offense. The male sample was around 3.5 times more likely than the female sample to have a Walsh conviction and also 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. Approximately 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 a work program. Males were about 4% more likely to complete at least one general program and 3% more likely to complete at least one work program than females. The one-, two-, and three-year general recidivism rates were 30.2%, 42.7%, and 48.5% for males and 18.1%, 26.9%, and 31.2% for females. The one-, two-, and three-year violent recidivism rates were 7.8%, 13.5%, and 18.3% for males and 2.3%, 4.1%, and 5.9% for females. The current sample is generally consistent with the prior release cohorts across these numerous metrics. (For comparison, see USDOJ 2021b, 14-15).

Table 1. ***Descriptive statistics for the FY 2018 sample, by gender***

Measure		% Total (*n* = 37,874)	% Male (*n* = 32,804)	% Female (*n* = 5,070)
Race/ethnicity	*White*	33.2	31.8	42.0
	Black	38.0	40.3	23.3
	Hispanic	22.9	22.2	27.3
	Asian	1.7	1.7	1.7
	Native American	4.2	4.0	5.7
Current age	*> 60*	4.7	4.8	4.0
	51-60	11.6	11.5	12.1
	41-50	22.1	22.1	22.3
	30-40	40.3	40.6	38.3
	26-29	12.5	12.3	13.6
	< 26	8.8	8.7	9.6

[10] An "act of violence" is defined based on the BOP Office of Research and Evaluation's (ORE) 19-category offense code classification. Six classifications are designated as violent: homicide, sexual assault, robbery, assault, weapons, and other violent. ORE's classification tracks statutory provisions related to violent offenses. See 18 U.S.C. § 16 and 18 U.S.C. § 3559(c)(2)(F); see also BOP Program Statement 5162.05, Categorization of Offenses.

Walsh with conviction	*No*	88.7	87.6	96.4
	Yes	11.3	12.4	3.6
Violent offense	*No*	70.1	66.9	90.8
	Yes	29.9	33.1	9.2
Criminal history points	*0-1*	28.5	25.5	47.8
	2-3	13.1	12.5	16.9
	4-6	17.6	17.9	15.8
	7-9	14.4	15.3	8.4
	10-12	10.4	11.2	4.9
	13+	16.0	17.5	6.2
History of escapes	*None*	82.9	82.1	88.3
	> 10 years minor	6.6	7.1	3.1
	5-10 years minor	3.6	3.7	3.2
	< 5 years minor or any serious	6.9	7.1	5.4
History of violence	*None*	47.9	43.8	74.8
	> 10 years minor	5.9	6.2	3.5
	> 15 years serious	9.6	10.8	1.9
	5-10 years minor	5.2	5.4	4.1
	10-15 years serious	6.8	7.6	1.8
	< 5 years minor	10.3	10.6	8.3
	5-10 years serious	9.3	10.3	3.3
	< 5 years serious	4.9	5.3	2.3
Education status	*Not enrolled*	19.8	19.2	23.2
	Enrolled in GED	10.5	10.4	11.1
	High school degree/GED	69.7	70.4	65.7
Drug program status	*Need indicated/no completion*	59.0	59.2	58.1
	Completed nonresidential drug treatment	5.9	5.9	5.8
	Completed residential drug treatment	5.2	5.1	5.7
	No need indicated	29.9	29.9	30.4
All incident reports	*0*	62.1	60.7	71.6
	1	16.3	16.6	14.4
	2	7.9	8.2	5.8
	3+	13.7	14.5	8.2
Serious incident reports	*0*	76.6	74.9	87.3
	1	13.6	14.4	8.4
	2	4.7	5.1	2.2
	3+	5.2	5.6	2.0
Time since last incident report	*12+ months*	75.7	75.1	79.6
	7-12 months	3.5	3.6	2.7
	3-6 months	4.7	4.9	3.6
	< 3 months	16.1	16.5	14.1
Time since last serious incident report	*12+ months*	86.9	86.0	92.2
	7-12 months	2.4	2.5	1.4
	3-6 months	3.1	3.3	1.8
	< 3 months	7.6	8.1	4.6
Financial responsibility refuse	*No*	96.0	95.9	96.3
	Yes	4.0	4.1	3.7
Programs completed	*0*	46.6	47.2	42.8
	1	15.9	15.7	17.2
	2-3	14.9	14.6	17.2
	4-10	16.8	16.6	18.2
	11+	5.8	6.0	4.7
Work programs completed	*0*	78.0	78.4	75.3
	1	14.0	13.6	17.2
	2+	78.0	8.0	7.5

General recidivism	*1 year*	28.5	30.2	18.1
	2 years	40.6	42.7	26.9
	3 years	46.2	48.5	31.2
Violent recidivism	*1 year*	7.1	7.8	2.3
	2 years	12.3	13.5	4.1
	3 years	16.7	18.3	5.9

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.[11] 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. Compared to previous years (USDOJ 2021b, 21), there are noticeable shifts in the general tool RLC distributions — reflections of the policy changes implemented by the Attorney General in April 2022 (USDOJ 2022). These redistributions caused more individuals to be placed into the lower RLG, making more individuals eligible for additional earned time credits under the FSA.

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

Risk assessment Risk level category (range of scores)	**First Assessment**	**Last Assessment**
Male general recidivism risk score (SD)	42.4 (20.3)	36.7 (22.1)
Percent minimum (-22 to 5)	3.9	9.8
Percent low (6 to 39)	38.8	42.2
Percent medium (40 to 54)	25.9	24.4
Percent high (55 to 109)	31.3	23.5
Male violent recidivism risk score (SD)	25.1 (12.8)	22.9 (14.0)
Percent minimum (-11 to 7)	8.3	15.3
Percent low (8 to 24)	40.5	39.0
Percent medium (25 to 31)	18.4	16.7
Percent high (32 to 71)	32.7	28.9
Female general recidivism risk score (SD)	25.6 (17.6)	19.5 (19.9)
Percent minimum (-27 to 7)	15.7	30.4
Percent low (8 to 38)	61.2	51.4
Percent medium (39 to 52)	15.4	12.4
Percent high (53 to 102)	7.7	5.8
Female violent recidivism risk score (SD)	4.4 (4.3)	2.2 (5.8)
Percent minimum (-11 to 1)	25.3	48.9
Percent low (2 to 11)	66.8	43.4
Percent medium (12 to 17)	7.3	6.6
Percent high (18 to 30)	0.5	1.1

[11] As part of the review process, and in partnership with BOP's Office of Research and Evaluation, the NIJ consultants assessed BOP's scoring and automation of PATTERN. They discovered one discrepancy involving the scoring of the education status measure. All four of the tools include the education status item, for which 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. Currently, individuals who are assigned by BOP for a GED program but not yet enrolled are classified by BOP as being enrolled. For the statistical modeling that was used to develop PATTERN version 1.3, however, the individuals assigned but not yet enrolled were not classified as participating. For the FY 2018 cohort, there were 3,181 males (9.7%) and 366 females (7.2%) designated as assigned but not yet enrolled.

Note: SD = standard deviation. Percentages do not all sum to 100 due to rounding. Male sample *n* = 32,804; female sample *n* = 5,070.

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 programming completed. As can be seen in Table 3, 88.6% of males (47.5% + 41.1%) and 87.0% of females (80.5% + 6.5%) 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	5,587 (47.5%)	2,240 (6.8%)
Violent higher RLG	1,486 (4.5%)	13,491 (41.1%)
Females	General lower RLG	General higher RLG
Violent lower RLG	4,084 (80.6%)	595 (11.7%)
Violent higher RLG	62 (1.2%)	329 (6.5%)

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* = 32,804; female sample *n* = 5,070.

Analytic Plan

The analytic plan corresponds with the approach taken in the prior USDOJ (2021b) report, with some additional analyses undertaken. As in the prior report, 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). The current report expands its predictive analyses by including 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. The report also provides 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.[12]

[12] As in previous years, 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

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.[13] Across all tools and follow-up periods, the AUC values were found to be strong predictors of recidivism (i.e., AUCs ≥ .714). Even the lower bounds of the 95% confidence intervals (CIs) were considered large in terms of predictive strength in 20 of the 24 analyses.[14] The AUCs were decisively higher at last assessment (range = .756 to .784) compared to first (range = .718 to .749).[15] For the two male 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 .756 at the one-year follow-up period, .774 at the two-year follow-up period, and .780 at the three-year follow-up period. In contrast, the AUCs for the two female tools were more variable across the different follow-up and assessment periods. Due to the lower number of women in the sample, the 95% CIs of the AUCs were also much wider for females relative to males, which suggests less precision in the confidence of the results.[16]

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		

indicated that for a given PATTERN score, members of tht 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 USDOJ 2021b, 37).

[13] As a supplemental analysis, study results were examined using only the 19,398 males (59.0% of the male sample) and the 4,489 females (88.5% 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 = .718 to .774), violent male (range = .729 to .763), general female (range = .727 to .777), and violent female samples (range = .708 to .764).

[14] This finding should be interpreted cautiously, as three of four 95% CIs with 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 = 5,070), but also violent recidivism among females was rare, with only 117 (or 2.3%) events at the one-year follow-up period, 209 (or 4.1%) at the two-year follow-up period, and 301 (or 5.9%) at the three-year follow-up period.

[15] 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.

[16] The differences detected between the male and female results are undoubtedly influenced by two factors. First, the size of the female sample (n = 5,070) was much smaller than that of the male sample (n = 32,874). And second, the recidivism rates of the female sample were much lower than those of the male sample. To illustrate, there were approximately 1.55 times more general recidivism events and 3.1 times more violent recidivism events for males relative to females at the three-year follow-up period. These findings do not invalidate the tool for females; rather, the results should be viewed as providing 95% confidence that the true AUC value for the general female tool at the three-year follow-up period is between .767 and .794, and that the AUC value for the female violent tool at the three-year follow-up period is between .741 and .792.

One-year follow-up	.718 [.712, .724]	.756 [.750, .761]
Two-year follow-up	.737 [.732, .742]	.774 [.768, .779]
Three-year follow-up	.746 [.741, .752]	.780 [.775, .785]
Male violent recidivism		
One-year follow-up	.741 [.732, .749]	.764 [.755, .772]
Two-year follow-up	.746 [.739, .753]	.768 [.762, .775]
Three-year follow-up	.749 [.742, .755]	.769 [.762, .775]
Female general recidivism		
One-year follow-up	.735 [.718, .752]	.772 [.756, .788]
Two-year follow-up	.747 [.733, .762]	.780 [.767, .794]
Three-year follow-up	.747 [.733, .761]	.779 [.766, .792]
Female violent recidivism		
One-year follow-up	.737 [.694, .781]	.784 [.743, .825]
Two-year follow-up	.740 [.708, .772]	.778 [.748, .809]
Three-year follow-up	.733 [.706, .760]	.767 [.741, .792]

Note: Male sample n = 32,804; female sample n = 5,070.

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, 10.5% of those rated as minimum-risk recidivated during the three-year follow-up period, followed by 32.7% in the low-risk group, 61.2% in the medium-risk group, and 79.6% in the high-risk group.

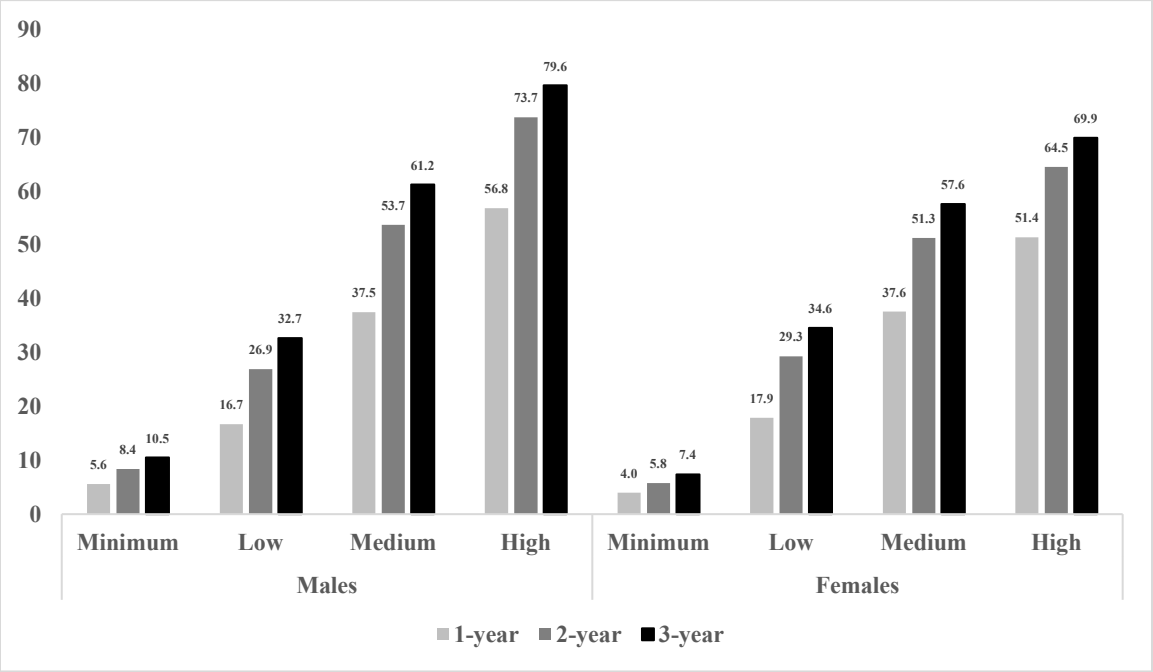

Figure 1. Percentage of males and females with general recidivism outcomes at one-, two-, and three-year follow-up periods by PATTERN last assessment general risk level category

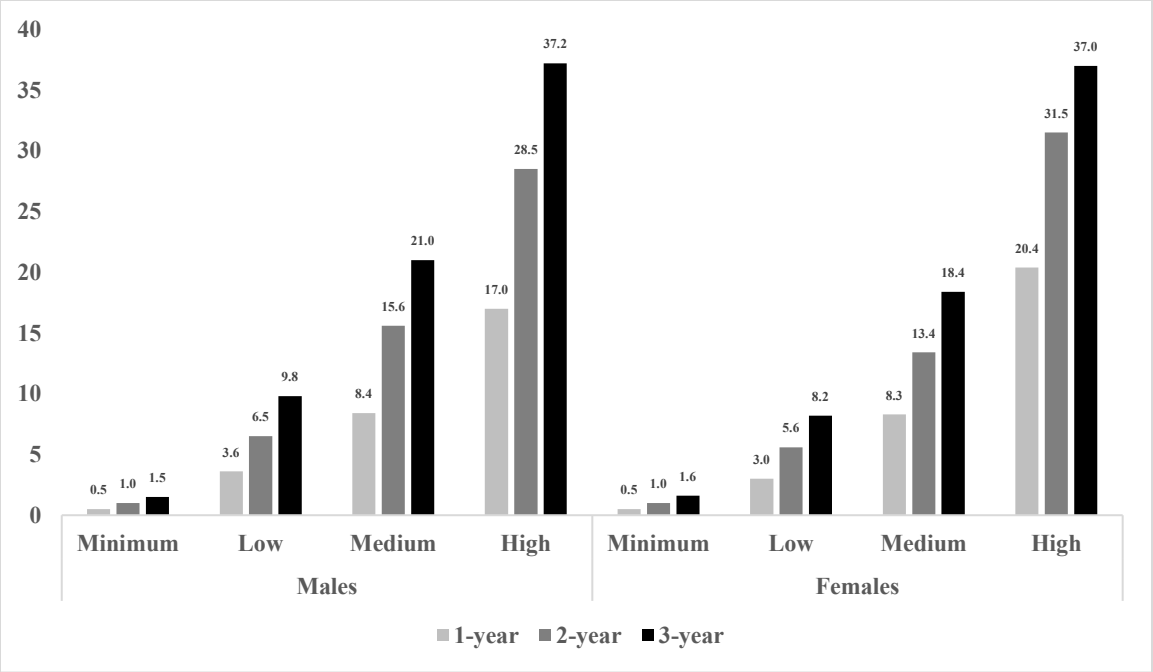

Figure 2. Percentage of males and females with violent recidivism outcomes at one-, two-, and three-year follow-up periods by PATTERN last assessment violent risk level category

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 directly on the RLGs. Consequently, the results for the general tools were impacted by the Attorney General's April 2022 adjustments to the cutpoints. Those values are thus not directly comparable to the values calculated for previous years, since the FY 2018 values reflect the new cutpoint implementation. Aside from the new cutpoints, however, the findings are similar to those from previous cohorts (see USDOJ 2021b, 25-26).

The PPV provides a measure of recidivism accuracy in the higher RLG, while the NPV provides a measure of nonrecidivism accuracy in the lower RLG. The higher RLG corresponds to recidivism events (PPVs) for 70% of males and 62% of females on the general tools, and 31% of males and 21% of females on the violent tools. For those classified in the lower RLG on the general tools, 71% of males and 76% of females avoided recidivism (NPVs). For the violent tools, the success rate was 93% for males and 95% for females. The full distribution tables are provided for one-, two-, and three-year outcomes 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	0.70	0.71	0.28	0.31
Male violent recidivism	0.31	0.93	0.38	0.22
Female general recidivism	0.62	0.76	0.10	0.64

Female violent recidivism	0.21	0.95	0.06	0.73

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 nonrecidivism); 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 = 32,804; female sample n = 5,070.

Part 4: Dynamic Validity

Table 6 presents changes in PATTERN risk scores and levels from first to last assessment.[17] 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 approximately an 8-point reduction detected for females and a 7-point reduction for males. 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 69%), approximately 24% to 33% had a lower risk level and 3% to 7% 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)	-6.8 (9.1)	-7.5 (8.7)
Percentage at lower risk level	28.8	29.1
Percentage at same risk level	67.8	67.2
Percentage at higher risk level	3.4	3.6
Violent recidivism		
Mean change in risk score (SD)	-2.6 (6.0)	-2.7 (3.7)
Percentage at lower risk level	24.2	32.9
Percentage at same risk level	69.2	62.3
Percentage at higher risk level	6.7	4.7

Note: SD = standard deviation. Percentages do not all sum to 100 due to rounding. Male sample n = 27,780; female sample n = 4,134.

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 -2.1 and -1.2 points, respectively. For the female sample, these differences were -1.2 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, 30.9% of the total change in score was due to change in score on the age variable (i.e., -2.1 points from age item divided by -6.8 points from total score). This same estimate was 46.2% for the male violent tool, 16.0% for the female general tool, and 7.4% 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

[17] The dynamic analyses included only the individuals with more than one assessment available. More specifically, there were 27,780 males (or 84.7% of the male sample) and 4,134 females (or 81.5% 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, 8.5% of males and 5.4% 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 (76.9% versus 41.5%, respectively). The proportion of females in the minimum-risk classification, however, roughly doubled between first and last assessment (from 16.6% to 34.5%).

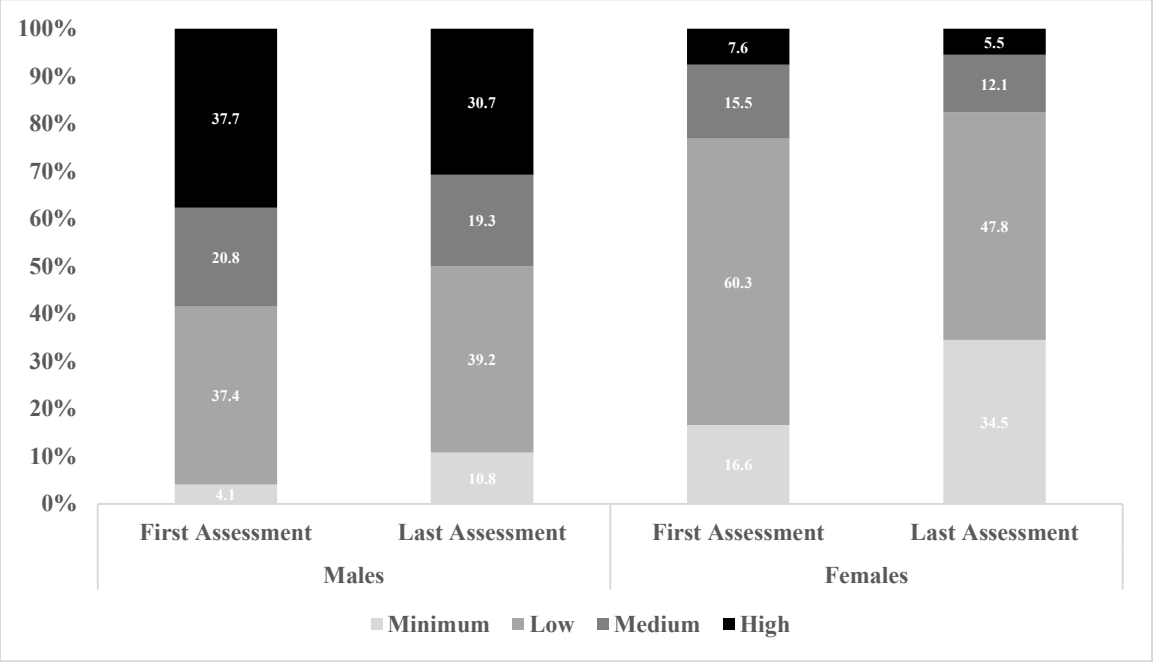

Figure 3. Percentage of males and females assigned to each of the overall PATTERN risk level categories at first and last assessment

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 all four instruments and 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, respectively. 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.

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	20.2	31.5	38.4
Same risk level	30.1	42.4	48.1
Higher risk level	47.2	62.8	68.6

Male Violent Recidivism			
Lower risk level	4.1	7.5	10.8
Same risk level	8.2	14.0	18.8
Higher risk level	10.0	17.2	22.9
Female General Recidivism			
Lower risk level	11.6	18.3	22.2
Same risk level	16.3	25.2	29.4
Higher risk level	35.6	47.0	52.3
Female Violent Recidivism			
Lower risk level	1.0	1.9	2.9
Same risk level	2.2	4.1	5.7
Higher risk level	7.2	12.3	16.4

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 = 27,780; Female sample n = 4,134.

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% to 8% 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 15% to 19% 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	1.06
Change in risk score	1.06	1.06	1.06
Constant	0.06	0.10	0.13
Model χ^2	5,065.7	6,765.5	7,318.2
Nagelkerke R^2	.240	.292	.309
Male Violent Recidivism			
First assessment score	1.08	1.08	1.09
Change in risk score	1.08	1.08	1.07
Constant	0.01	0.02	0.02
Model χ^2	1,918.7	3,112.5	3,985.5
Nagelkerke R^2	.164	.199	.223
Female General Recidivism			
First assessment score	1.05	1.05	1.06
Change in risk score	1.07	1.07	1.06
Constant	0.06	0.10	0.13
Model χ^2	558.8	764.1	854.8
Nagelkerke R^2	.218	.253	.269
Female Violent Recidivism			
First assessment score	1.19	1.19	1.18

Change in risk score	1.19	1.17	1.15
Constant	0.01	0.02	0.03
Model χ^2	110.2	170.9	202.7
Nagelkerke R^2	.145	.148	.141

Note: Reported values are odds ratios. All findings are statistically significant at the .001 level. Male sample n = 27,780; Female sample n = 4,134.

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, 11.5% of white males are age 29 or younger, while 21.7% of Black males and 32.4% of Hispanic males are age 29 or younger. There are differences in criminal history across groups as well. Over 30% of white, Hispanic, and Asian males fall into the lowest criminal history point category (0-1), while only 12.7% of Black males fall into that category. Similar but less pronounced age differences appear for the females; for example, 17.7% of white females are age 29 or younger, while 21.3% of Black females are age 29 or younger. Native American and Black individuals 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 (53.5% and 72.7%) and violent recidivism (24.2% and 26.0%), while Asian males had the lowest rates of these two outcomes (31.8% and 11.3%). The female recidivism rates exhibited a notably different trend: Native American and white females had the highest rates of observed three-year general recidivism (56.8% and 33.4%), and Native American and Black females had the highest rates of observed three-year violent recidivism (12.5% and 6.8%). Asian females had the lowest rates of both general and violent recidivism (15.9% and 3.4%). It is also important to point out that Black and Hispanic females both had lower rates of general recidivism (26.8% and 27.4%) than white females (33.4%).

Table 9. ***Descriptive statistics for the FY 2018 male release sample, by race/ethnicity***

Measure		% White (*n* = 10,427)	% Black (*n* = 13,229)	% Hispanic (*n* = 7,288)	% Asian (*n* = 551)	% Native American (*n* = 1,309)
Current age	*> 60*	9.6	2.7	2.2	7.3	1.9
	51-60	18.0	8.8	7.5	15.2	8.2
	41-50	25.3	21.7	18.5	28.9	17.6
	30-40	35.6	45.1	39.4	37.4	41.9
	26-29	7.8	13.4	16.4	7.1	16.1
	< 26	3.7	8.3	16.0	4.2	14.3
Walsh with conviction	*No*	79.7	91.4	94.8	91.8	68.5
	Yes	20.3	8.6	5.2	8.2	31.5
Violent offense	*No*	72.9	57.9	80.3	80.4	31.2
	Yes	27.1	42.1	19.7	19.6	68.8
Criminal history points	*0-1*	33.3	12.7	36.8	46.8	20.9
	2-3	12.7	9.6	15.4	15.2	22.4
	4-6	15.8	19.9	17.0	14.3	22.1
	7-9	12.2	19.3	12.6	9.4	16.5
	10-12	8.8	15.3	8.1	4.9	9.5
	13+	17.1	23.3	10.0	9.3	8.7
History of escapes	*None*	83.6	79.0	87.8	92.6	64.6
	> 10 years minor	7.0	9.1	3.8	3.1	8.5
	5-10 years minor	3.0	4.6	2.4	2.2	7.3
	< 5 years minor or any serious	6.3	7.3	5.9	2.2	19.6
History of violence	*None*	52.2	32.4	55.6	60.4	18.6
	> 10 years minor	6.7	7.2	4.4	5.1	2.9
	> 15 years serious	9.3	15.1	6.2	6.7	6.3
	5-10 years minor	4.8	6.2	4.8	4.5	5.3
	10-15 years serious	6.2	9.7	5.7	8.3	8.3
	< 5 years minor	9.3	11.4	10.7	5.3	15.1
	5-10 years serious	7.8	12.2	8.4	6.7	22.2
	< 5 years serious	3.5	5.7	4.3	2.9	21.2
Education status	*Not enrolled*	12.5	18.5	28.8	18.3	27.3
	Enrolled in GED	5.0	12.5	14.4	7.1	11.5
	High school degree/GED	82.5	68.9	56.8	74.6	61.2
Drug program status	*Need indicated/no completion*	54.1	59.4	62.9	45.7	82.3
	Completed nonresidential drug treatment	5.9	6.1	5.9	4.9	4.1
	Completed residential drug treatment	5.5	5.1	5.0	5.3	2.1
	No need indicated	34.5	29.5	26.2	44.1	11.5
All incident reports	*0*	66.3	56.4	58.9	73.0	63.9
	1	15.1	17.4	17.9	13.6	14.5

	2	7.0	9.4	8.3	5.1	7.0
	3+	11.6	16.8	14.9	8.3	14.7
Serious incident reports	*0*	79.5	71.4	73.7	85.8	75.2
	1	11.6	16.2	15.3	9.3	14.4
	2	3.7	6.1	5.4	2.0	4.9
	3+	5.1	6.2	5.6	2.9	5.4
Time since last incident report	*12+ months*	78.8	73.4	73.0	83.7	70.4
	7-12 months	3.1	4.0	3.7	2.9	3.1
	3-6 months	4.0	5.4	5.2	2.2	6.6
	< 3 months	14.1	17.2	18.2	11.3	20.0
Time since last serious incident report	*12+ months*	88.2	85.2	84.8	92.7	81.3
	7-12 months	2.3	2.6	2.7	1.5	2.4
	3-6 months	2.8	3.6	3.6	1.3	4.9
	< 3 months	6.7	8.6	8.9	4.5	11.4
Financial responsibility refuse	*No*	97.2	95.7	94.9	98.2	92.6
	Yes	2.8	4.3	5.1	1.8	7.4
Programs completed	*0*	46.5	45.3	48.5	41.2	66.7
	1	16.3	15.6	15.2	18.3	12.4
	2-3	14.8	14.7	15.0	14.0	9.2
	4-10	15.9	18.3	15.9	19.1	9.2
	11+	6.6	6.1	5.4	7.4	2.6
Work programs completed	*0*	78.5	76.7	79.6	77.7	87.9
	1	14.0	14.0	13.1	12.9	8.5
	2+	7.5	9.3	7.3	9.4	3.6
General recidivism	*1 year*	28.9	31.8	25.7	18.7	52.7
	2 years	39.5	46.3	37.0	28.3	68.4
	3 years	44.5	53.5	42.1	31.8	72.7
Violent recidivism	*1 year*	5.6	10.6	5.6	5.3	11.5
	2 years	9.5	18.2	10.0	8.2	20.4
	3 years	13.5	24.2	13.8	11.3	26.0

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

Table 10. ***Descriptive statistics for the FY 2018 female release sample, by race/ethnicity***

Measure		% White (*n* = 2,129)	% Black (*n* = 1,181)	% Hispanic (*n* = 1,385)	% Asian (*n* = 88)	% Native American (*n* = 287)
Current age	*> 60*	5.1	4.1	2.7	8.0	1.0
	51-60	15.3	11.3	8.4	17.0	7.7
	41-50	23.9	25.1	17.8	29.5	18.8
	30-40	38.0	38.1	37.8	33.0	44.9
	26-29	12.4	13.2	15.8	8.0	15.7
	< 26	5.3	8.1	17.5	4.5	11.8
Walsh with conviction	*No*	96.5	94.2	98.2	93.2	96.5
	Yes	3.5	5.8	1.8	6.8	3.5
Violent offense	*No*	92.2	88.7	96.3	94.3	61.7
	Yes	7.8	11.3	3.7	5.7	38.3
Criminal history points	*0-1*	39.4	50.4	60.6	64.8	32.4
	2-3	17.9	14.5	14.5	14.8	31.7
	4-6	19.5	14.1	11.9	8.0	15.7
	7-9	10.2	7.3	5.7	4.5	13.9
	10-12	5.7	4.9	4.0	4.5	3.8
	13+	7.2	8.9	3.2	3.4	2.4
History of escapes	*None*	88.7	87.0	90.8	93.2	78.0
	> 10 years minor	3.7	4.1	1.7	1.1	2.4
	5-10 years minor	3.2	3.6	2.2	3.4	5.2
	< 5 years minor or any serious	4.4	5.4	5.3	2.3	14.3
History of violence	*None*	77.5	66.5	81.4	84.1	54.7
	> 10 years minor	4.1	4.5	1.9	1.1	3.5
	> 15 years serious	1.9	3.5	0.9	1.1	1.0
	5-10 years minor	4.6	3.9	3.0	6.8	5.9
	10-15 years serious	1.9	2.2	1.3	2.3	0.7
	< 5 years minor	6.5	10.7	7.9	2.3	15.3
	5-10 years serious	2.3	5.5	1.9	1.1	9.1
	< 5 years serious	1.2	3.3	1.6	1.1	9.8
Education status	*Not enrolled*	18.1	21.9	31.2	21.6	28.6
	Enrolled in GED	6.4	11.8	17.0	9.1	15.7
	High school degree/GED	75.5	66.3	51.8	69.3	55.7
Drug program status	*Need indicated/no completion*	60.2	52.2	56.3	39.8	80.5
	Completed nonresidential drug treatment	7.5	5.6	3.9	2.3	4.9
	Completed residential drug treatment	7.4	5.2	3.7	9.1	3.5
	No need indicated	24.9	37.0	36.1	48.9	11.1
All incident reports	*0*	75.2	65.6	70.7	81.8	71.8
	1	13.8	15.4	14.9	9.1	14.6

	2	4.6	8.2	6.1	4.5	3.1
	3+	6.5	10.8	8.4	4.5	10.5
Serious incident reports	*0*	88.5	84.2	88.2	92.0	85.7
	1	7.9	9.8	8.2	4.5	9.1
	2	1.7	3.0	2.2	3.4	2.4
	3+	1.8	3.0	1.4	0.0	2.8
Time since last incident report	*12+ months*	83.0	75.4	78.1	85.2	77.4
	7-12 months	2.2	4.1	2.6	1.1	1.4
	3-6 months	3.0	4.1	4.0	2.3	3.8
	< 3 months	11.8	16.4	15.3	11.4	17.4
Time since last serious incident report	*12+ months*	93.0	90.5	92.3	94.3	91.3
	7-12 months	1.4	1.5	1.4	0.0	1.4
	3-6 months	1.4	2.2	2.0	2.3	2.8
	< 3 months	4.2	5.8	4.3	3.4	4.5
Financial responsibility refuse	*No*	97.4	94.9	96.1	97.7	95.1
	Yes	2.6	5.1	3.9	2.3	4.9
Programs completed	*0*	41.1	39.8	44.5	40.9	60.3
	1	20.7	15.8	14.1	19.3	10.8
	2-3	16.7	16.9	18.4	18.2	15.7
	4-10	17.0	20.6	19.6	17.0	10.8
	11+	4.5	6.9	3.5	4.5	2.4
Work programs completed	*0*	75.3	76.8	73.5	72.7	79.1
	1	16.1	16.6	19.1	20.5	17.4
	2+	8.6	6.6	7.4	6.8	3.5
General recidivism	*1 year*	18.6	16.7	15.7	8.0	34.8
	2 years	28.7	23.3	23.1	13.6	51.9
	3 years	33.4	26.8	27.4	15.9	56.8
Violent recidivism	*1 year*	1.5	3.6	1.9	1.1	5.2
	2 years	3.2	4.9	4.0	2.3	8.7
	3 years	4.6	6.8	6.1	3.4	12.5

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 (43.1 and 45.5 points) than white, Hispanic, and Asian males (30.0, 34.0, 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 (59.8% and 66.3%) than white, Hispanic, and Asian males (38.2%, 39.0%, and 23.1%).[18] Native American females had much higher general recidivism scores at final assessment (27.8 points) than white, Black, Hispanic, and Asian females, who all had similar general recidivism scores (19.3, 19.5, 18.6, and 9.9 points). Native American females were also more likely to be classified in the higher RLG of the general recidivism tool (26.9%) than white, Black, Hispanic, and Asian females (18.5%, 20.1%, 15.2%, and 7.9%).

[18] 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***

	White		Black		Hispanic		Asian		Native American	
Risk assessment	**First**	**Last**	**First**	**Last**	**First**	**Last**	**First**	**Last**	**First**	**Last**
Male general recidivism										
Mean risk score (SD)	35.9 (22.1)	30.0 (24.2)	49.5 (17.8)	43.1 (19.5)	38.9 (17.9)	34.0 (20.2)	29.6 (21.1)	22.9 (21.6)	47.7 (16.2)	45.5 (18.4)
Percent minimum	8.6	18.9	1.0	3.5	2.3	8.5	10.0	23.0	1.1	3.1
Percent low	46.9	42.8	26.1	36.8	51.2	52.4	58.1	53.9	26.5	30.6
Percent medium	20.4	19.4	30.0	29.4	25.4	22.0	15.8	12.0	37.3	32.8
Percent high	24.2	18.8	42.9	30.4	21.1	17.0	16.2	11.1	35.1	33.5
Male violent recidivism										
Mean risk score (SD)	20.8 (13.3)	18.4 (14.8)	29.4 (11.7)	26.8 (12.7)	22.9 (11.2)	21.3 (12.6)	17.7 (13.0)	14.6 (13.4)	32.4 (10.8)	32.3 (12.2)
Percent minimum	17.0	27.6	2.7	6.5	6.1	14.2	21.8	36.1	2.4	4.0
Percent low	44.0	37.5	32.0	37.2	54.1	47.8	47.9	39.7	19.9	20.8
Percent medium	15.4	13.5	21.8	20.1	16.5	15.5	13.2	10.0	21.4	18.2
Percent high	23.6	21.4	43.5	36.2	23.3	22.5	17.1	14.2	56.3	57.1
Female general recidivism										
Mean risk score (SD)	26.2 (18.1)	19.3 (20.1)	26.0 (18.5)	19.5 (21.2)	23.9 (15.9)	18.6 (18.6)	16.5 (18.0)	9.9 (18.4)	31.0 (14.5)	27.8 (17.5)
Percent minimum	17.1	29.9	15.9	33.5	14.0	30.4	40.9	54.5	5.9	14.3
Percent low	57.9	51.7	59.1	46.4	68.2	54.4	44.3	37.5	65.5	58.9
Percent medium	16.8	13.3	14.9	11.9	12.3	10.4	13.6	6.8	22.0	19.2
Percent high	8.3	5.2	10.1	8.2	5.4	4.8	1.1	1.1	6.6	7.7
Female violent recidivism										
Mean risk score (SD)	4.3 (4.1)	1.8 (5.4)	4.7 (4.8)	2.5 (6.4)	3.8 (3.8)	1.9 (5.3)	2.3 (4.1)	-0.2 (4.9)	7.5 (5.0)	6.4 (6.3)
Percent minimum	23.8	49.8	28.5	50.6	27.2	50.0	46.6	71.5	7.7	22.6
Percent low	70.2	44.7	60.1	38.7	68.1	44.8	52.3	26.1	67.9	52.6
Percent medium	5.8	5.1	10.2	8.5	4.5	4.6	1.1	2.3	22.0	21.6
Percent high	0.2	0.4	1.2	2.3	0.1	0.6	0.0	0.0	2.4	3.1

Note: SD = standard deviation. Percentages do not all sum to 100 due to rounding. White male sample *n* = 10,427; Black male sample *n* = 13,229; Hispanic male sample *n* = 7,288; Asian male sample *n* = 551; Native American male sample *n* = 1,309; white female sample *n* = 2,129; Black female sample *n* = 1,181; Hispanic female sample *n* = 1,385; Asian female sample *n* = 88; Native American female sample *n* = 287.

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 ≥ .714) in 93 of the 120 analyses (or 77.5%). The remaining AUC values were all considered moderate predictors of recidivism (i.e., AUCs ≥ .639). 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 .695 (Native American female violent recidivism) to a high of .904 (Asian female violent recidivism).[19] Additionally, the AUCs were higher across all four tools and three follow-up periods at last assessment compared to first among all five of the racial and ethnic groups. The results with the three-year outcomes are consistent with those reported with the prior release cohorts (see USDOJ 2021b, 34).

[19] 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***

	One-year follow-up		Two-year follow-up		Three-year follow-up	
Risk assessment	**First**	**Last**	**First**	**Last**	**First**	**Last**
Male general recidivism						
White	.750 [.740, .759]	.777 [.767, .786]	.767 [.757, .776]	.791 [.783, .800]	.775 [.766, .784]	.798 [.789, .806]
Black	.690 [.681, .700]	.740 [.731, .749]	.707 [.698, .715]	.757 [.749, .765]	.716 [.707, .725]	.763 [.755, .771]
Hispanic	.725 [.712, .738]	.756 [.744, .768]	.734 [.722, .746]	.767 [.756, .778]	.733 [.722, .745]	.763 [.752, .774]
Asian	.804 [.761, .847]	.836 [.796, .876]	.812 [.774, .849]	.842 [.807, .877]	.801 [.763, .838]	.828 [.792, .863]
Native American	.670 [.641, .700]	.680 [.651, .708]	.714 [.684, .745]	.719 [.688, .749]	.717 [.684, .750]	.720 [.688, .753]
Male violent recidivism						
White	.773 [.756, .790]	.788 [.772, .805]	.773 [.760, .787]	.790 [.777, .803]	.780 [.768, .791]	.794 [.783, .805]
Black	.700 [.686, .713]	.736 [.723, .749]	.705 [.694, .716]	.742 [.732, .752]	.708 [.698, .717]	.742 [.732, .751]
Hispanic	.731 [.707, .754]	.753 [.732, .775]	.731 [.713, .749]	.752 [.735, .769]	.731 [.715, .747]	.750 [.735, .766]
Asian	.833 [.785, .880]	.845 [.800, .889]	.832 [.789, .876]	.838 [.792, .885]	.810 [.763, .857]	.813 [.763, .863]
Native American	.669 [.625, .713]	.686 [.645, .728]	.691 [.657, .726]	.695 [.661, .728]	.692 [.660, .723]	.697 [.666, .728]
Female general recidivism						
White	.696 [.670, .722]	.741 [.716, .766]	.723 [.700, .745]	.764 [.743, .785]	.727 [.705, .748]	.765 [.744, .785]
Black	.768 [.733, .804]	.794 [.759, .828]	.770 [.739, .801]	.797 [.768, .827]	.762 [.732, .792]	.791 [.762, .819]
Hispanic	.761 [.728, .794]	.791 [.760, .822]	.769 [.742, .797]	.791 [.765, .818]	.766 [.739, .793]	.791 [.765, .816]
Asian	.837 [.659, 1.00]	.886 [.749, 1.00]	.825 [.698, .951]	.853 [.741, .964]	.825 [.713, .938]	.864 [.767, .962]
Native American	.683 [.620, .746]	.728 [.670, .787]	.663 [.600, .726]	.689 [.628, .750]	.663 [.598, .727]	.696 [.634, .758]
Female violent recidivism						
White	.646 [.555, .737]	.701 [.617, .785]	.704 [.648, .760]	.743 [.689, .797]	.707 [.660, .754]	.727 [.680, .774]
Black	.769 [.700, .838]	.804 [.733, .874]	.786 [.729, .844]	.817 [.760, .873]	.787 [.738, .836]	.817 [.771, .864]
Hispanic	.782 [.709, .855]	.825 [.762, .889]	.728 [.666, .791]	.774 [.716, .832]	.717 [.664, .770]	.757 [.708, .806]
Asian	.966 [.919, 1.00]	.989 [.966, 1.00]	.837 [.647, 1.00]	.869 [.694, 1.00]	.835 [.702, .968]	.904 [.778, 1.00]
Native American	.657 [.499, .815]	.716 [.580, .853]	.687 [.572, .802]	.721 [.619, .823]	.649 [.547, .752]	.695 [.605, .785]

Note: White male sample n = 10,427; Black male sample n = 13,229; Hispanic male sample n = 7,288; Asian male sample n = 551; Native American male sample n = 1,309; white female sample n = 2,129; Black female sample n = 1,181; Hispanic female sample n = 1,385; Asian female sample n = 88; Native American female sample n = 287.

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 82% of the time; there was some variation across racial and ethnic groups, with the higher RLG designation being less accurate for Hispanics at 66% and most accurate for Native American males at 82%. Overall, the violent male PPV was 31%, reflecting the greater difficulty in predicting lower prevalence events (see USDOJ 2021b). Here, the tool was most accurate for Black males at 35% and least accurate for Hispanic males at 25%. For female general recidivism the overall PPV was 62%, with white, Black, and Hispanic females having PPVs at 60% or 61%. Native American females had the highest PPVs at 74% and Asian females the lowest at 57%.

Turning to the NPVs, lower risk males did not recidivate 46%-79% of the time, with the rate for Native American males at 46% representing an outlier. This finding suggests that PATTERN may be less effective at correctly identifying lower risk Native American males, as this race 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 were quite accurate, with NPVs ranging from 90% to 95% for males (being lowest for Black males at 90%). For females the violent RLGs were accurate 95% of the time, ranging from a low of 92% for Native American females to a high of 98% for Asian females.

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

Risk assessment tool	*PPV*	*NPV*	*FPR*	*FNR*
Male general recidivism				
White	0.72	0.73	0.19	0.38
Black	0.70	0.70	0.39	0.22
Hispanic	0.66	0.73	0.23	0.39
Native American	0.82	0.46	0.43	0.25
Asian	0.69	0.79	0.11	0.50
Total	0.70	0.71	0.28	0.31
Male violent recidivism				
White	0.28	0.94	0.29	0.27
Black	0.35	0.90	0.48	0.18
Hispanic	0.25	0.93	0.33	0.31
Native American[a]	0.32	0.93	0.69	0.07
Asian[a]	0.31	0.95	0.19	0.34
Total	0.31	0.93	0.38	0.22
Female general recidivism				
White	0.60	0.73	0.11	0.67
Black	0.61	0.82	0.11	0.54
Hispanic	0.60	0.79	0.08	0.66
Native American[a]	0.74	0.50	0.16	0.65
Asian[a]	0.57	0.88	0.04	0.71
Total	0.62	0.76	0.10	0.64
Female violent recidivism				
White[a]	0.12	0.96	0.05	0.86
Black	0.28	0.96	0.08	0.55
Hispanic[a]	0.18	0.95	0.05	0.85
Native American[a]	0.25	0.92	0.21	0.50

Asian[a]	0.50	0.98	0.01	0.67
Total	0.21	0.95	0.06	0.73

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* = 10,427; Black male sample *n* = 13,229; Hispanic male sample *n* = 7,288; Asian male sample *n* = 551; Native American male sample *n* = 1,309; white female sample *n* = 2,129; Black female sample *n* = 1,181; Hispanic female sample *n* = 1,385; Asian female sample *n* = 88; Native American female sample *n* = 287.

Table 14 reports the differential prediction analyses. These results are based on the PATTERN score, not the RLCs, so they are not affected by the 2022 cutpoint changes (USDOJ 2022). The results (Table 13 and Figure 4) show that the Black and Hispanic male relative overprediction documented in USDOJ (2021b) continued to worsen for FY 2018, with Black males now about 5% less likely to engage in general recidivism compared to white males (and Hispanic males about 6% less likely) when holding the PATTERN score constant. As in previous years, Native American risk was significantly underpredicted compared to white males on the general recidivism tool (see USDOJ 2021b, 36). The violent male risk tool exhibited results that were mostly not statistically significant, with the exception being a 3.7% 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).

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) at 6.9%, 4.6%, and 9.9%, respectively. Native American female general risk was underpredicted by about 13.4% compared to white females, also consistent with previous years. Finally, female violent results were mostly not statistically significant and similar to previous years. The Model 4 interaction effects were mostly not statistically significant. One exception was on the general female tool. As the differential prediction plot in Figure 3 shows, Hispanic females with lower PATTERN scores were overpredicted relative to white females, as they were less likely to recidivate given the same PATTERN score. However, Hispanic females with higher PATTERN scores were underpredicted relative to white females. The plot lines appear to cross around a score in the low 40s, and the Table 13 results confirm a statistically significant result. (The visualization from the plot suggests a similar trend for Asian females, but those results were not statistically significant.) Another exception was with Black males on the general tool, but there the difference was mostly relative overprediction throughout the scores, with slight underprediction at the very highest score values. Although these results were statistically significant, we note that the confidence intervals (not plotted for clarity) at the extreme values of the PATTERN score are the largest, reflecting the sparse distribution of individuals in those extreme values.[20]

[20] The differential prediction results were 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. For example, the Black-to-white relative overprediction margins were: -6.9% (Black male, general, one-year), -6.1% (Black male, general, two-year), -2.5% (Black female, general, one-year), and -5.8% (Black female, general, two-year).

Table 14. ***Summary of differential prediction regression analyses findings***

	Model 3				Model 4			
General Male	**AME**	**OR**	***p***		**AME**	**OR**	***p***	
Black	-4.97%	0.77	0.000	***	-4.99%	0.63	0.000	***
Hispanic	-5.98%	0.73	0.000	***	-5.82%	0.69	0.000	***
Native American	11.36%	1.85	0.000	***	12.29%	2.11	0.000	***
Asian	-4.95%	0.77	0.015	*	-2.95%	0.53	0.003	**
Black × score						1.01	0.002	**
Hispanic × score						1.00	0.365	
Native American × score						1.00	0.501	
Asian × score						1.01	0.042	*
Violent Male								
Black	3.73%	1.33	0.000	***	3.68%	1.42	0.001	***
Hispanic	-0.94%	0.92	0.095		-1.02%	1.01	0.923	
Native American	-0.14%	0.99	0.874		0.50%	1.47	0.136	
Asian	1.35%	1.12	0.461		1.77%	1.01	0.982	
Black × score						1.00	0.506	
Hispanic × score						1.00	0.418	
Native American × score						0.99	0.107	
Asian × score						1.00	0.699	
General Female								
Black	-6.89%	0.66	0.000	***	-6.90%	0.63	0.003	**
Hispanic	-4.62%	0.76	0.001	***	-4.43%	0.56	0.000	***
Native American	13.35%	2.03	0.000	***	14.67%	2.39	0.001	***
Asian	-9.90%	0.54	0.058		-7.24%	0.23	0.030	*
Black × score						1.00	0.615	
Hispanic × score						1.01	0.016	*
Native American × score						0.99	0.501	
Asian × score						1.04	0.112	
Violent Female								
Black	1.00%	1.23	0.209		0.94%	0.89	0.635	
Hispanic	1.64%	1.38	0.042	*	1.73%	1.23	0.353	
Native American	2.11%	1.49	0.070		3.04%	1.85	0.111	
Asian	0.18%	1.04	0.949		2.54%	0.35	0.392	
Black × score						1.05	0.089	
Hispanic × score						1.02	0.465	
Native American × score						0.99	0.711	
Asian × score						1.21	0.181	

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 = 10,427; Black male sample n = 13,229; Hispanic male sample n = 7,288; Asian male sample n = 551; Native American male sample n = 1,309; white female sample n = 2,129; Black female sample n = 1,181; Hispanic female sample n = 1,385; Asian female sample n = 88; Native American female sample n = 287.

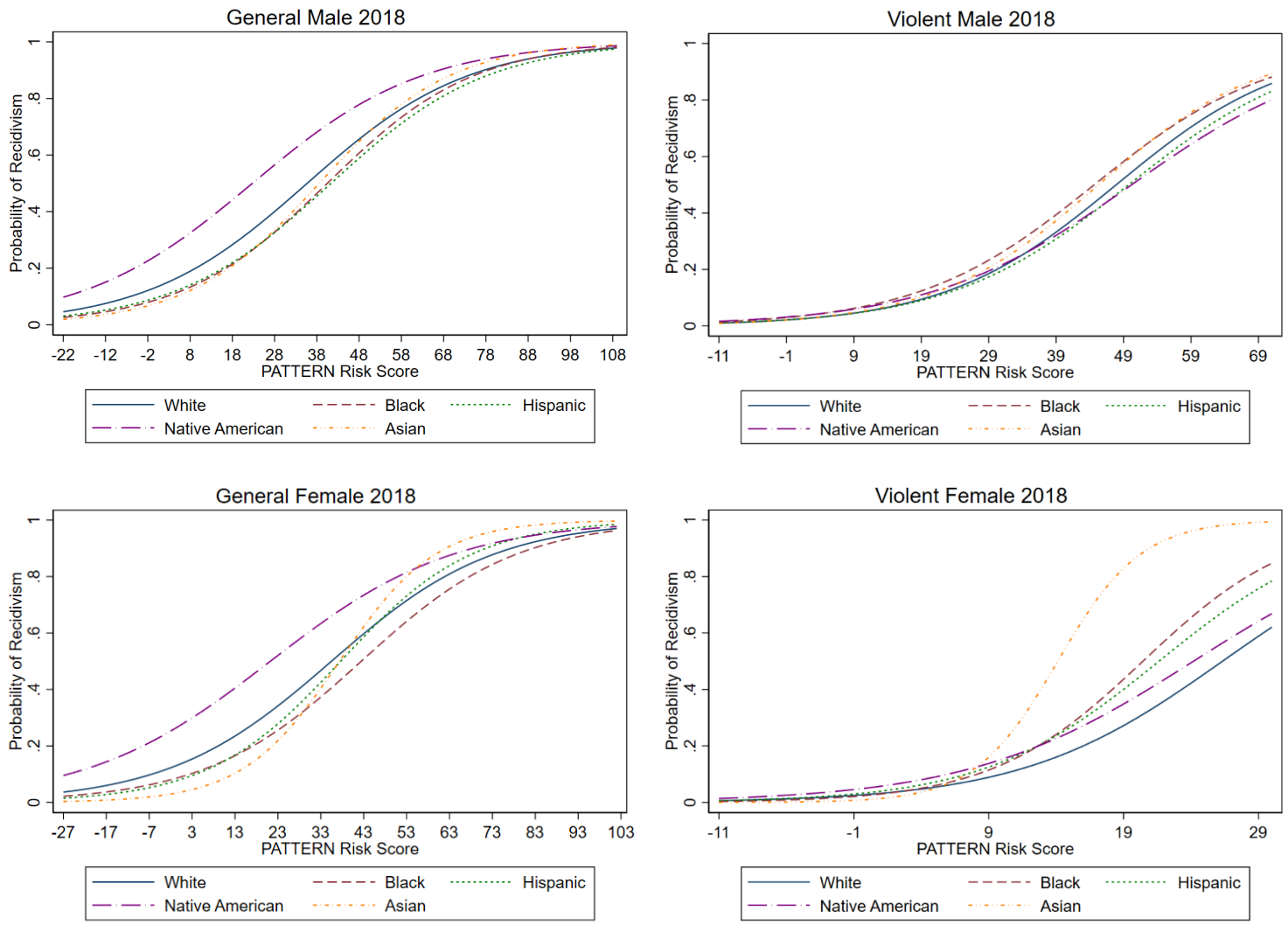

Figure 4. Differential prediction plots, by PATTERN instrument

Part 6: Discussion and Conclusion

This study represents the third 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. In addition to replicating the analyses undertaken on prior cohorts with a subsequent cohort of FY 2018 BOP releasees, the current study expanded its analyses to include one-, two-, and three-year follow-up recidivism outcomes, additional analyses on what accounts for change in risk scores and levels, and additional descriptive information on individual items, risk scores and levels, and outcomes by racial and ethnic group.

The current study findings continue to demonstrate that PATTERN is a strong predictor of general and violent recidivism, with AUCs ranging from 0.76 to 0.78. The tools predict both types of recidivism at the one-, two-, and three-year follow-up periods. 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 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).

The findings also document meaningful differences across the RLC-by-race distributions. Although racial and ethnic neutrality can be examined through numerous metrics,[21] 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.[22]

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) report. That report 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,

[21] 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."

[22] 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).

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 part 1, last year's review and revalidation report acknowledged "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" (USDOJ 2021b, 46). The report stated that NIJ and its consultants would "continue to investigate potential solutions for the differential prediction issues identified" (USDOJ 2021b, 47). Efforts to address these issues to date are detailed in part 1 and include the Attorney General's 2022 changes to the RLCs to mitigate racial and ethnic impact, and ongoing efforts to reconsider the data and methodology based on engagement with experts and stakeholders. NIJ and DOJ are committed to these efforts to fulfill the FSA mandate to pursue potential revisions or updates to PATTERN "to ensure that any disparities identified … are reduced to the greatest extent possible" (FSA § 3631 (b)(5)).

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 (BOP). (2006). *Inmate security designation and custody classification* (policy number 5100.08). U.S. Department of Justice.

Federal Bureau of Prisons (BOP). (2021). *Evidence-based recidivism reduction (EBRR) programs and productive activities (PA)*. U.S. Department of Justice.

Federal Bureau of Prisons (BOP). (2022a, March). *First Step Act: Initial review of the SPARC-13 needs assessment system*. U.S. Department of Justice.

Federal Bureau of Prisons (BOP). (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). (2022, April). *First Step Act annual report*. Office of the Attorney General.

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: *25 and younger*, *26 to 29*, *30 to 40*, *41 to 50*, *51 to 60*, or *61 and older*.
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 BOP 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: *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*.
History of escapes	Years since last escape attempt by seriousness (from the most recent BRAVO); converted into four ordinal categories: *None*, *>10 years minor*, *5 to 10 years minor*, or *less than 5 years minor or any serious*.
History of violence	Number of years from last act of violence by seriousness (from the most recent BRAVO); converted into eight ordinal categories: *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*.
Education status	The highest grade level completed; converted into three ordinal categories: *High school degree or GED*, *enrolled and progressing in GED program*, or *no verified degree and not participating in GED program*.
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 nonresidential drug programming during the current incarceration.[a] This variable is then converted into four ordinal categories: *No drug need indicated*, *completed residential drug treatment*, *completed nonresidential drug treatment*, or *need indicated but no drug treatment completed*.
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 BOP incarcerations; converted into four ordinal categories: *No incident*, *1 incident*, *2 incidents*, or *3 or more incidents*.
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 BOP incarcerations; converted into four ordinal categories: *No incident*, *1 incident*, *2 incidents*, or *3 or more incidents*.
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: *12+ months or no incident*, *7 to 12 months*, *3 to 6 months*, or *less than 3 months*.
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: *12+ months or no incident*, *7 to 12 months*, *3 to 6 months*, or *less than 3 months*.
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: *No program*, *1 program*, *2 to 3 programs*, *4 to 10 programs*, or *11 or more programs*.[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: *No program*, *1 program*, or *2 or more programs*.

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 BOP, as these data were not available during the study observation period (see BOP 2021).

[b] This measure does not include all the EBRR programs and PAs currently available in BOP, 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 BOP policy, though they predictively correlate with recidivism (see BOP 2021).

Appendix B

PATTERN Scoring Guide

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

12. Time since last serious incident report	*12+ months*			0	0
	7-12 months			1	1
	3-6 months			2	2
	< 3 months			3	3
13. Financial responsibility refuse	*No*	0		0	0
	Yes	2		3	1
14. Programs completed	*0*	0	0	0	0
	1	-3	-1	-2	-1
	2-3	-6	-2	-4	-2
	4-10	-9	-3	-6	-3
	11+	-12	-4	-8	-4
15. Work programs completed	*0*	0	0	0	0
	1	-1	-1	-2	-1
	2+	-2	-2	-4	-2
Risk level category	*Minimum*	-22 to 5	-11 to 7	-27 to 7	-11 to 1
	Low	6 to 39	8 to 24	8 to 38	2 to 11
	Medium	40 to 54	25 to 31	39 to 52	12 to 17
	High	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	.718 [.710, .726]	.762 [.754, .769]
Two-year follow-up	.728 [.721, .735]	.771 [.764, .777]
Three-year follow-up	.734 [.727, .741]	.774 [.767, .780]
Male violent recidivism		
One-year follow-up	.729 [.715, .744]	.762 [.748, .775]
Two-year follow-up	.734 [.723, .745]	.763 [.753, .774]
Three-year follow-up	.731 [.721, .741]	.758 [.748, .767]
Female general recidivism		
One-year follow-up	.727 [.709, .746]	.765 [.747, .783]
Two-year follow-up	.746 [.731, .762]	.777 [.762, .792]
Three-year follow-up	.745 [.730, .760]	.776 [.762, .791]
Female violent recidivism		
One-year follow-up	.708 [.655, .761]	.764 [.714, .814]
Two-year follow-up	.719 [.680, .757]	.761 [.724, .798]
Three-year follow-up	.712 [.680, .744]	.748 [.718, .779]

Note: Male sample n = 19,398; female sample n = 4,489.

Appendix D

Risk level categories and recidivism by race and ethnicity, FY 2018 revalidation sample, one-year follow-up period

	Male General		Male Violent		Female General		Female Violent	
	Did Not Recidivate	*Recidivated*	*Did Not Recidivate*	*Recidivated*	*Did Not Recidivate*	*Recidivated*	*Did Not Recidivate*	*Recidivated*
White								
Minimum	1846	127	2868	12	606	30	1054	6
Low	3631	834	3781	124	881	219	929	22
Medium	1173	852	1303	107	185	98	105	4
High	765	1199	1894	338	62	48	9	0
Black								
Minimum	449	15	858	8	378	18	593	4
Low	4153	713	4706	209	475	73	439	18
Medium	2566	1317	2405	250	85	55	87	13
High	1849	2167	3862	931	46	51	20	7
Hispanic								
Minimum	586	33	1027	5	409	12	690	2
Low	3241	580	3378	109	636	118	601	19
Medium	1025	581	1054	76	93	51	60	4
High	560	682	1423	216	29	37	7	2
Native American								
Minimum	37	4	52	0	40	1	64	1
Low	248	152	262	10	115	54	145	6
Medium	197	233	217	21	25	30	56	6
High	137	301	627	120	7	15	7	2
Asian								
Minimum	124	3	199	0	47	1	63	0
Low	258	39	210	9	30	3	23	0
Medium	43	23	47	8	4	2	1	1
High	23	38	66	12	0	1	0	0

Risk level categories and recidivism by race and ethnicity, FY 2018 revalidation sample, two-year follow-up period

	Male General		Male Violent		Female General		Female Violent	
	Did Not Recidivate	*Recidivated*	*Did Not Recidivate*	*Recidivated*	*Did Not Recidivate*	*Recidivated*	*Did Not Recidivate*	*Recidivated*
White								
Minimum	1783	190	2856	24	594	42	1049	11
Low	3179	1286	3687	218	742	358	904	47
Medium	870	1155	1207	203	139	144	98	11
High	479	1485	1691	541	44	66	9	0
Black								
Minimum	440	24	851	15	373	23	592	5
Low	3635	1231	4535	380	429	119	432	25
Medium	1916	1967	2197	458	69	71	83	17
High	1107	2909	3234	1559	35	62	16	11
Hispanic								
Minimum	572	47	1024	8	403	18	685	7
Low	2880	941	3280	207	565	189	581	39
Medium	776	830	994	136	77	67	57	7
High	362	880	1262	377	20	46	7	2
Native American								
Minimum	35	6	52	0	36	5	64	1
Low	193	207	254	18	78	91	140	11
Medium	114	316	193	45	18	37	53	9
High	72	366	543	204	6	16	5	4
Asian								
Minimum	123	4	197	2	46	2	63	0
Low	230	67	208	11	27	6	22	1
Medium	31	35	42	13	3	3	1	1
High	11	50	59	19	0	1	0	0

Risk level categories and recidivism by race and ethnicity, FY 2018 revalidation sample, three-year follow-up period

	Male General		Male Violent		Female General		Female Violent	
	Did Not Recidivate	*Recidivated*	*Did Not Recidivate*	*Recidivated*	*Did Not Recidivate*	*Recidivated*	*Did Not Recidivate*	*Recidivated*
White								
Minimum	1743	230	2843	37	583	53	1043	17
Low	2937	1528	3562	343	678	422	885	66
Medium	745	1280	1135	275	118	165	97	12
High	361	1603	1481	751	39	71	7	2
Black								
Minimum	433	31	847	19	367	29	589	8
Low	3313	1553	4349	566	405	143	421	36
Medium	1564	2319	2036	619	61	79	75	25
High	838	3178	2801	1992	31	66	16	11
Hispanic								
Minimum	556	63	1020	12	397	24	680	12
Low	2679	1142	3183	304	526	228	560	60
Medium	681	925	946	184	69	75	53	11
High	301	941	1132	507	14	52	7	2
Native American								
Minimum	34	7	52	0	35	6	63	2
Low	170	230	248	24	69	100	135	16
Medium	91	339	181	57	15	40	49	13
High	62	376	488	259	5	17	4	5
Asian								
Minimum	120	7	194	5	46	2	63	0
Low	216	81	203	16	25	8	21	2
Medium	30	36	38	17	3	3	1	1
High	10	51	54	24	0	1	0	0