

Written Testimony of

Melissa Hamilton, JD, Ph.D
Professor of Law & Criminal Justice
University of Surrey School of Law
England

Before the

United States House of Representatives
House Committee on the Judiciary
Subcommittee on Crime, Terrorism, and Homeland Security

Regarding

“The First Step Act, The Pandemic, and Compassionate Release:
What Are the Next Steps for the Federal Bureau of Prisons?”

January 21, 2022

Esteemed members: Thank you for your invitation to appear before you today.

I address herein the Prisoner Assessment Tool Targeting Risk and Needs (PATTERN), the risk and needs system that was created under the auspices of the First Step Act of 2018. The development and implementation of a risk tool with a large, national population is a tremendously difficult endeavor. In such a process, there are inevitably errors, disagreements, controversies, and trade-offs to be made.

As an academic, I do not advocate for or against the use of risk assessment tools in criminal justice settings. That choice is for policymakers with the difficult responsibility of making those decisions. Instead, the purpose is to highlight some pressing issues that PATTERN presents to inform policymakers, other stakeholders, and the public. It is noted first that transparency regarding PATTERN has waxed and waned. This could be due to issues of tight time deadlines and the pandemic. Notably, the most recent NIJ Review and Revalidation report of December 2021¹ provides a healthy amount of information. I will refer to it here as the NIJ Report. Many concerns, though, remain to be highlighted and resolved.

The Evolution of PATTERN

PATTERN evaluates males and females separately.² For each gender, there is a general recidivism scale (i.e., any rearrest) and a violent recidivism scale (i.e., any violent rearrest). Fifteen risk factors have been identified, though not all of them are used in each of the four scales (i.e., the four are male general, male violent, female general, and female violent). For each of the general and violent recidivism scales, scores are combined into four ordinal risk categories of minimum, low, medium, and high risk. The greater of the risk categories becomes the individual's final risk category. For instance, if an individual is classified as medium risk in the general scale and low risk in the violence scale, the individual is assigned a final category of medium risk.

PATTERN has undergone various iterations. Using the numbering system employed in the NIJ Report, four versions have existed:

- The initial draft of PATTERN was publicly released in July 2019.³
- PATTERN 1.2 was a revision following receipt and consideration of comments from listening sessions with stakeholders and solicitations of public feedback. PATTERN 1.2 was approved for use by the Attorney General and has been used to score individuals in Bureau of Prisons (BOP) custody from January 2020 onward.
- PATTERN 1.2-R corrected for "typos" identified in the BOP scoring sheets so that the tool in practice was the version approved for use.⁴ The risk level categories currently assigned to prisoners in the BOP are based on PATTERN 1.2-R.

¹ National Institute of Justice, 2021 Review and Revalidation of the First Step Act Risk Assessment Tool (December 2021), <https://nij.ojp.gov/library/publications/2021-review-and-revalidation-first-step-act-risk-assessment-tool> [hereinafter NIJ Report 2021].

² There are substantive, legitimate reasons (scientifically and legally) to differentiate by gender as studies consistently indicate that the risk profiles and likelihood of recidivism vary significantly for males versus females.

³ Department of Justice, The First Step Act of 2018: Risk and Needs Assessment System (2019), https://nij.ojp.gov/sites/g/files/xyckuh171/files/media/document/the-first-step-act-of-2018-risk-and-needs-assessment-system_1.pdf.

⁴ NIJ Report 2021, at 7.

- PATTERN 1.3 is a version offered in the December 2021 NIJ Report that (i) corrects for additional errors the newly installed consultants discovered existing within PATTERN 1.2 (and 1.2-R) and (ii) rehails certain of the risk factors, weights, scoring sheets, definitions, and sources of data from which to draw. PATTERN 1.3 is not in use as the NIJ reported that details about it were issued to begin discussions with the Independent Review Committee and others, with the potential of later being proposed to the Attorney General for approval.⁵ Only at that time will the BOP implement an update in lieu of PATTERN 1.2-R.

Errors in PATTERN

At present, the BOP assigns prisoners the risk category levels produced by PATTERN 1.2-R. As indicated above, the reason given is that it represents the version approved to date by the Attorney General. Yet, the NIJ Report makes clear that this version (1.2-R) contains many errors and that, as a result, 10.9% of males and 9.8% of females have been given incorrect overall PATTERN risk categories as a result.⁶ In other terms, as of late 2020, a total of 14,170 prisoners have been assigned erroneous final risk categories.⁷ The BOP has no plans to correct these errors until a new version of PATTERN (such as the proposed version 1.3) is formally approved by the Attorney General.⁸

A summary of these errors from NIJ publications include these:

- Four errors in scoring or cut-point: (1) point values for the infraction-free (serious and violent) item for the violent male scale, (2) point values for the infraction-free (serious and violent) item for the general recidivism female scale, (3) point values for the criminal history score for the violent recidivism female scale, (4) the cut-points for the general recidivism male scale.⁹
- The initial version of the tool was developed based on a statistical model that scored a vast majority of the risk factors *at the time of one's release from custody* rather than what was specified in the formal publications which designed these factors to be scored *at the time of the individual's last assessment* (which typically would occur some period of time before release). As a simple example of why this may matter, consider a risk factor regarding the individual's age. One's age at the time of release will be older than at the time of an in-custody assessment, and this difference may have changed the scoring on this age-related risk factor. An earlier (published in 2020) NIJ publication admits the broader issue: "Because the empirical models were estimated using different versions of these variables, it may have influenced the coefficients obtained and the item weights assigned."¹⁰ In other words, this definitional discrepancy across risk factors called into question the efficacy of the entire scoring system.
- There were irregularities in the tool's intended function and what occurred in practice in scoring the number of prison disciplinary infractions the individual had (impacting four risk factors). The

⁵ NIJ Report 2021, at 47.

⁶ NIJ Report 2021, at 7.

⁷ National Institute of Justice, 2020 Review and Revalidation of the First Step Act Risk Assessment Tool 7-8 tbls. 1-2 (January 2021), <https://www.ojp.gov/pdffiles1/nij/256084.pdf> [hereinafter NIJ Review 2020].

⁸ A group of 1,745 individuals whose scores were impacted by the scoring "typos" corrected in PATTERN 1.2-R were rescored and their risk levels updated. NIJ Review 2020, at 10 n. 39.

⁹ NIJ Review 2020, at 5.

¹⁰ NIJ Review 2020, at 6.

difference was whether to count infractions no matter how old or only those occurring within the last 10 years. Then BOP personnel also counted infractions when individuals were in pretrial and holdover stages, whereas PATTERN 1.2 would exclude infractions during those types of custody. The NIJ sums up the problem: “This means that as BOP is implementing PATTERN 1.2, they are currently scoring these infraction variables differently than were modeled in the reported PATTERN 1.2, which may have an impact on the utility of these two measures.”¹¹

- The developmental study improperly defined several risk factors: (i) the number of programs completed was inadvertently counted as the number of programs in which the person participated (regardless of completion), (ii) counted participation in UNICOR as a multiplier of the number of days in UNICOR rather than treating participation in UNICOR as one program regardless of number of days, and (iii) counted participation in drug treatment while imprisoned as having completed only a nonresidential drug treatment program for individuals who had completed both residential and nonresidential programs. As per an NIJ publication the potential impact is this: “Given that the empirical models were estimated using different constructs of the variables, the factor selection and weighting may have resulted in a scoring and risk assessment scheme that would be different if the correct versions of these variables were included.”¹²

As a consequence of such errors, officials called off the initial revalidation of PATTERN.¹³ A revalidation was eventually conducted, with findings published in the (December 2021) NIJ Report. The results indicated that the various errors meant that 37 out of the possible 60 items (almost two-thirds of them) had been incorrectly weighted.¹⁴ Due to these errors, according to the NIJ Report, overall, 11% of the BOP population was placed in the wrong risk category. This proportion may be on the low end. The NIJ Report also indicates a significant problem with reliability in that BOP personnel incorrectly scored and classified more than 20% of the BOP population.¹⁵ An automated system has been developed to improve reliability. However, it is unclear when/if the misclassifications from manual scoring will be remedied.

In sum, as the flawed PATTERN 1.2-R continues to be used, erroneous risk level classifications appear to be continuing to this day.

An Issue with Revalidation

The (December 2021) NIJ Report labels itself a “Revalidation of the First Step Act Risk Assessment Tool.” This terminology is curious considering the differences between the versions of PATTERN: the one that has been and is currently used in practice to score individual risk levels (version 1.2-R) and the modified version proposed for adoption (version 1.3). While such Report provides some information on the empirical performance of PATTERN 1.2-R, the vast majority of the statistical measures provided are for

¹¹ NIJ Review 2020, at 6.

¹² NIJ Review 2020, at 6.

¹³ NIJ Review 2020, at 6. Other problems plagued the developmental study. For example, the initial publication indicated that individuals who had died after release and before the end of the follow-up period were excluded, but in fact they were not. NIJ Review 2020, at 3 n.9.

¹⁴ There are 15 possible risk factors and four scales (male general, male violence, female general, female violence) and thus up to 60 (15 x 4) items.

¹⁵ NIJ Report 2021, at 8.

PATTERN 1.3. Hence, this NIJ Report in reality seems more of a validation of the new PATTERN version 1.3.

Preference for False Positives

PATTERN operates with significant rates of error and disproportionately prefers false positives over false negatives. A false positive is the incorrect prediction of higher risk (i.e., a person classified as high risk is not rearrested), whereas a false negative is the incorrect prediction of lower risk (i.e., a person classified as low risk was rearrested). This means that a choice has been made to design the tool to perform far less accurately when predicting those who are at higher risk—which means placing too many individuals into the higher risk groupings than necessary.

The preference for a high proportion of false positives is not a necessary one dictated by any scientific principles. As the First Step Act was designed, there is little danger to the public of incorrect predictions as the risk assessment outcomes are not meant to lead to immediate release. Indeed, even with earned time credits, the individual's sentence is not shortened. Instead, the predictions relate to who is given more robust incentives to engage with rehabilitative programming and who might earn a change in the type of prerelease custody. Thus, a policy directive could be given to recalibrate PATTERN to reduce the numbers of false positives, which in turn would increase the number of individuals who are eligible to work toward earned time credits.

Racial and Ethnic Differences

The NIJ Report informs that PATTERN does not perform equally based on race and ethnicity. The new NIJ consultants deserve praise for helpfully providing multiple metrics and for showing results across various groupings. Still, using the conclusions of the Report itself, the tool overpredicts the general risk for African Americans, Hispanic Americans, and Asian Americans, while it underpredicts for Native Americans. An explanation given for the underprediction of Native Americans is the lack of information to score this group on criminal history and on recidivism considering tribal reservations generally have not been required to provide arrest information to the National Law Enforcement Telecommunications System.¹⁶

There are also differences in the rates at which PATTERN assigns individuals to risk classifications based on race, as indicated in Table 1, and by ethnicity in Table 2 (utilizing abbreviated race/ethnic labels assigned by the DOJ in the publication from which these statistics are derived).¹⁷ Note that Table 1 includes within each racial category those who are also identified as Hispanic (or not) (e.g., white Hispanic, black Hispanic).

¹⁶ NIJ Report 2021, at 43 n. 55.

¹⁷ Statistics obtained from Department of Justice, Federal Prisoner Statistics Collected under the First Step Act, 2021, at 16 tbl. 10 (November 2021, NCJ 301582), <https://bjs.ojp.gov/content/pub/pdf/fpscfsa21.pdf>. The Asian category also includes Native Hawaiian and other Pacific Islander. The American Indian category also includes Alaska Native.

Table 1

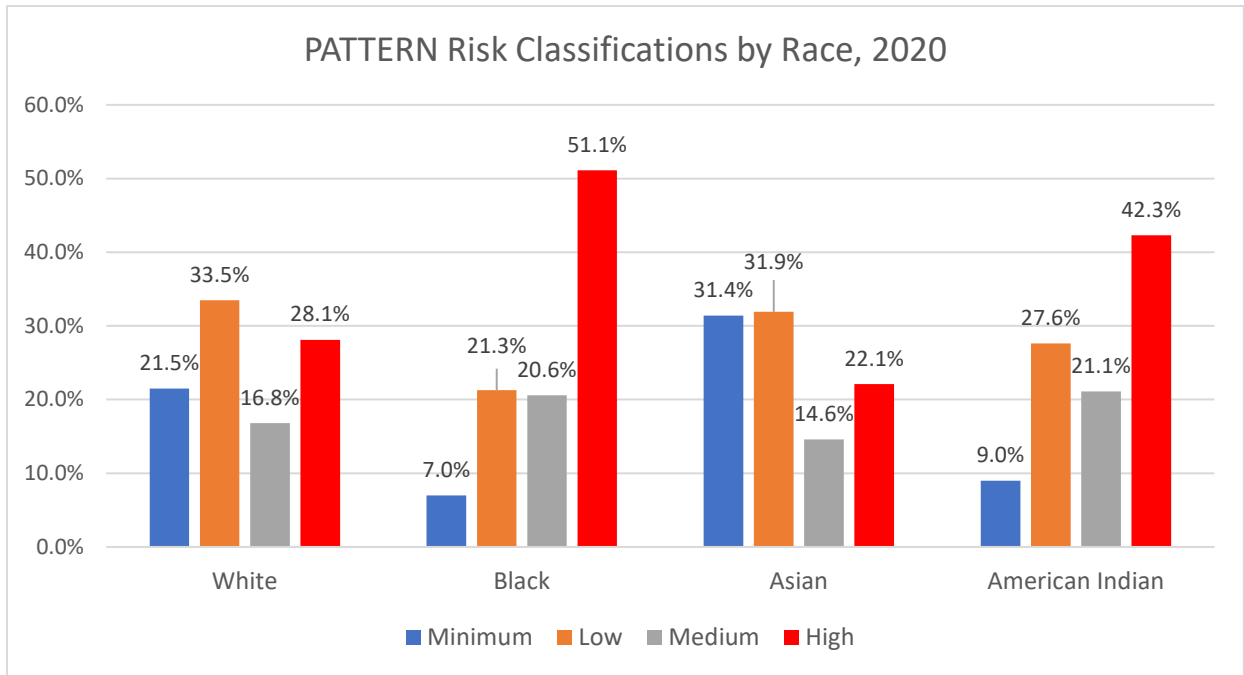
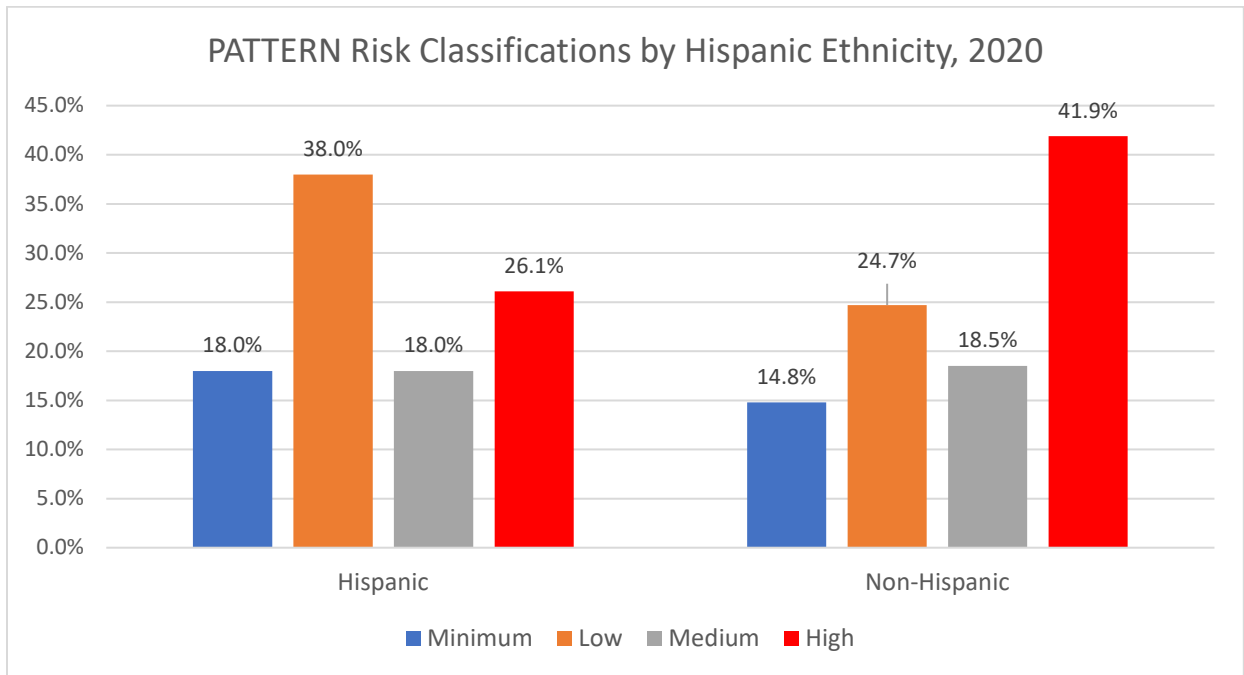


Table 1 shows that PATTERN classifies African Americans as high risk at a disproportionate rate (51%) compared to other groups, such as 28% of Whites and 22% of Asians.

Table 2



Concerning potential racial/ethnic differences overall, I concur with these comments in the NIJ Report: “multiple definitions of racial fairness exist, and that in real-world applications, these notions of fairness

conflict” and “a tool cannot satisfy all definitions of fairness.”¹⁸ When recidivism rates vary across groups, one or more of the standards for racial fairness are likely to be violated. Nonetheless, further work can be done toward ameliorating these potential disparities, which the NIJ Report suggests.

PATTERN Risk Levels and Compassionate Release

The deployment of PATTERN as an informational point in determining who to release for compassionate reasons and/or COVID represents an off-label use. The tool was not designed or validated for such a purpose.

Needs System

PATTERN is not itself a needs system.¹⁹ Instead, the BOP is relying, and purportedly improving, upon its preexisting policies and practices of identifying individual needs.²⁰ This means that to date there has been no (publicly known) validation of the needs aspect of the broader system. The BOP states that it is working to identify appropriate programs. At this time, though, a significant divide exists between program availability and individual demand in many BOP facilities.²¹ The result is a sort of lottery system whereby the luck of the draw in facility placement means some individuals will have a greater access to achieving earned time credits than others.

Release of Datasets for Independent Evaluation

The NIJ Report is helpful in providing a host of various statistics to provide outsiders a better understanding of PATTERN. However, this is not a fully acceptable alternative to making publicly available an anonymized version of the dataset(s) for independent researchers. I have discovered various statistical and textual errors in the NIJ Report itself. As with the revelation of problems in the initial PATTERN development by new consultants, verification of the work of these consultants might well be better confirmed by others.

Final Thoughts

In conclusion, I remain hopeful that there is a path for Congressional intent to be realized with the First Step Act. This will require continued efforts to correct the current inaccurate ratings, brainstorm on ways to reduce disparities, supplement the availability of programs, and to validate the needs component.

¹⁸ NIJ Report 2021, at 44.

¹⁹ Department of Justice, The Attorney General’s First Step Act Section 3634 Annual Report (December 2020), https://www.bop.gov/inmates/fsa/docs/20201221_fsa_section_3634_report.pdf.

²⁰ Department of Justice, The Attorney General’s First Step Act Section 3634 Annual Report, at Section II (December 2020), https://www.bop.gov/inmates/fsa/docs/20201221_fsa_section_3634_report.pdf.

²¹ Department of Justice, The Attorney General’s First Step Act Section 3634 Annual Report 17-18 tb. 11 (December 2020), https://www.bop.gov/inmates/fsa/docs/20201221_fsa_section_3634_report.pdf.