REVALIDATION OF THE FEDERAL PTRA

Testing the PTRA for Predictive Biases

THOMAS H. COHEN

CHRISTOPHER T. LOWENKAMP
Administrative Office of the U.S. Courts

The pretrial risk assessment instrument (PTRA) was developed for use in the U.S. federal pretrial system. Specifically, this instrument was constructed to help officers assess the likelihood that defendants will commit pretrial violations including being rearrested for new crimes, missing court appearances, or being revoked while on pretrial release. This research evaluates the PTRA's capacity to predict pretrial violations on 85,369 defendants with officer-completed PTRA assessments. Bivariate and multivariate models were estimated by race, ethnicity, and sex. Results show that the PTRA performs well at predicting pretrial violations as the area under the receiver operating characteristic curve ranged from 0.65 to 0.73 depending upon the subsamples and outcomes being predicted. Moreover, the PTRA predicted new criminal arrest activity equally well for non-Hispanic Whites and Blacks, while for Hispanics and females, findings show the instrument validly predicting rearrest activity, with some evidence of overprediction depending upon the outcome being examined.

Keywords: pretrial risk assessment instrument; federal pretrial release; recidivism prediction; predictive bias

When a person is arrested and accused of a crime in the federal system, a judicial official must determine whether the accused person (i.e., the defendant) will be released back into the community or detained until their case is disposed (American Bar Association, 2007). The decision to release or detain a defendant pretrial represents one of the most crucial components of the criminal justice process (Eskridge, 1983; Goldkamp, 1985). In addition to curtailing a defendant's liberty and constitutional rights, the decision to detain a defendant pretrial can potentially affect case outcomes by increasing the likelihood of conviction as well as the length of an imposed sentence (Cohen & Reaves, 2007; Harrington & Spohn, 2007; Hart & Reaves, 1999; Heaton, Mayson, & Stevenson, 2017; Lowenkamp,

VanNostrand, & Holsinger, 2013; Oleson, VanNostrand, Lowenkamp, Cadigan, & Wooldredge, 2014; Stevenson & Mayson, 2017; Tartaro & Sedelmaier, 2009; Ulmer, 2012). Several rationales have been provided as possible explanations for the adverse effects of pretrial detention on case outcomes, including the incentivization of pleading guilty to get out of jail, the inability to prepare an adequate defense or forestall the prosecution while incarcerated, and the diminishment of the capacity to engage in model positive behaviors while on release (Heaton et al., 2017). Furthermore, evidence suggests that defendants detained pretrial are more likely to commit future crimes than their similarly situated released counterparts (Gupta, Hansman, & Frenchman, 2016; Heaton et al., 2017; Lowenkamp et al., 2013; Oleson et al., 2014).

Given the importance of the pretrial release decision on case outcomes and subsequent recidivism activity, the process is increasingly being informed by actuarial risk instruments capable of assessing a defendant's risk of committing pretrial violations involving missed court appearances or threats to public safety (Bechtel, Lowenkamp, & Holsinger, 2011; Criminal Justice Policy Program, 2016; Mamalian, 2011; Pepin, 2013). Actuarial methods differ from clinical approaches where decision makers rely on professional judgment or intuition gleaned through interviews or documentation reviews to best assess offender risk (Andrews & Bonta, 2010; Connolly, 2003; Van Voorhis & Brown, 1996). Actuarial tools assess risk by statistically measuring the importance of various factors shown by the literature to be empirically related to recidivism (Andrews & Bonta, 2010). Most of these tools operate through checklist methods in which those factors significantly correlated with recidivism are assigned specific scores which are then summed to generate an overall risk score, although recently more complex machine learning approaches have been developed to assess recidivism risk (Mayson, 2018).

The federal pretrial system has embraced the use of actuarial decision-making techniques by adopting a risk assessment instrument titled the pretrial risk assessment instrument (hereafter, PTRA) to assess a defendant's likelihood of engaging in pretrial violations involving threats to public safety or missed court appearances (Cadigan, Johnson, & Lowenkamp, 2012; Cadigan & Lowenkamp, 2011; Lowenkamp & Whetzel, 2009). Implemented in November 2009, the PTRA has nearly universal usage rates. As the PTRA is being extensively used in the federal pretrial system, ongoing research is required to ensure its validity. Moreover, it is important to examine whether characteristics such as race moderate the relationship between the PTRA and the various pretrial outcomes it is purported to predict.

While several scholars have highlighted the potential for decisions based on risk assessments to produce disparate impacts for persons of color such as Blacks by unintentionally placing them into risk categories higher than their actual behavior justifies (i.e., the problem of false positives and negatives; see Chouldechova, 2016; Chouldechova & G'Sell, 2017; Hamilton, 2015; Harcourt, 2015; Kleinberg, Mullainathan, & Raghavan, 2016; Klingele, 2016; Silver & Miller, 2002; Starr, 2014), the key issue we address involves that of calibration. Calibration refers to whether the instrument predicts pretrial violation outcomes equally well regardless of whether a defendant is White, Black, Hispanic, male, or female (Skeem & Lowenkamp, 2016; Skeem, Monahan, & Lowenkamp, 2016). Hence, the current research will endeavor to accomplish two key objectives. Initially, it will seek to revalidate the PTRA on a large national sample of released federal defendants with actual PTRA assessments. Next, this research will investigate the PTRA for predictive biases by gauging

whether this instrument predicts pretrial rearrests equally well for Blacks and Hispanics compared with non-Hispanic Whites and males compared with females (Gottfredson & Gottfredson, 1988; Hoge, 2002; Skeem & Lowenkamp, 2016). Before delving into these issues, a brief overview of the PTRA is provided for background purposes. Included will be a discussion of the importance of focusing on the PTRA's predictive calibration by race, Hispanic ethnicity, and sex. Afterward, study methods will be detailed and principal findings presented. The study will conclude by discussing implications for the risk assessment and pretrial fields and elaborate directions for future research.

THE FEDERAL PTRA TOOL

The development and implementation of the PTRA in the federal pretrial system is well documented (see Cadigan et al., 2012; Cadigan & Lowenkamp, 2011; Lowenkamp & Whetzel, 2009). In brief, construction and validation samples comprising about 200,000 federal defendants released pretrial between fiscal years 2001 and 2007 were used to construct a risk instrument capable of predicting a released defendant's risk of failure to appear (FTA; e.g., missed court appearances), rearrests for new criminal activity, or pretrial revocation (Cadigan et al., 2012; Lowenkamp & Whetzel, 2009). In addition to predicting these events separately, the PTRA was constructed to predict combinations of these outcomes, including any adverse events (i.e., violations involving a combination of pretrial revocations, new criminal rearrests, or missed court appearances), or a combined new criminal rearrest/FTA outcome.

Using logistic regression modeling techniques, 11 items were identified and incorporated into the PTRA's risk-scoring algorithm (Cadigan et al., 2012; Lowenkamp & Whetzel, 2009). These items include factors measuring a defendant's criminal history, instant conviction offense, age, educational attainment, employment status, residential ownership, substance abuse problems, and citizenship status. For a detailed overview of the PTRA's development and the items and scores associated with this instrument, see Cadigan et al. (2012) and Lowenkamp and Whetzel (2009). It should be noted that many of these items are used by other pretrial risk assessments (see Bechtel et al., 2011; Laura and John Arnold Foundation [LJAF], 2013). Weights for these items were calculated based on the magnitude of the bivariate relationship between the selected factors and the pretrial violation outcomes mentioned above and ranged from 0 to 3 points, depending upon the item being scored. Ultimately, this process resulted in a risk-scoring algorithm that generated raw scores for each defendant ranging from 0 to 15 that were further grouped through visual inspection and confirmation of best fit as determined by chi-square analysis into the following five risk categories: PTRA 1 (scores 0-4), PTRA 2 (scores 5-6), PTRA 3 (scores 7-8), PTRA 4 (scores 9-10), or PTRA 5 (scores 11 or above; Lowenkamp & Whetzel, 2009). Both the initial validation and revalidation studies showed the PTRA successfully differentiating defendants by their risk of garnering pretrial violations involving FTA, new criminal rearrests, and pretrial revocations (Cadigan & Lowenkamp, 2011; Lowenkamp & Whetzel, 2009).

While these studies show the PTRA serving as an adequate predictive mechanism, as is the case with any risk assessment, ongoing validation is required, as is investigating the instrument's validity with subpopulations of interest. The last PTRA revalidation occurred several years ago and was done on a small sample of released federal defendants ($n = 5,077$) with actual officer-completed PTRA assessments (Cadigan et al., 2012). In addition,

to date, there has been no published research on the predictive validity of the PTRA for subpopulations based on whether the defendant is White, Black, Hispanic, male, or female.[1] We did not seek to validate the PTRA on other non-White populations (e.g., Asians, Pacific Islanders, American Indians, or Alaska Natives), as these groups constituted relatively small portions of the federal pretrial population. Thus, the issue of calibration (meaning whether the PTRA predicts equally well across race, Hispanic ethnicity, and sex) is of paramount importance and will be investigated in depth in this research study (Corbett-Davies, Pierson, Feller, Goel, & Huq, 2017).

THE PTRA AND PREDICTIVE BIAS

Rather than focusing on whether risk instruments predict outcomes equally well across race, Hispanic ethnicity, and sex subpopulations, several scholars assert that the items integrated into risk instruments, including those measuring socioeconomic characteristics such as marital history, employment status, educational attainment (see Starr, 2014, 2015), or criminal history (see Harcourt, 2008), might serve as proxies for race (Skeem & Lowenkamp, 2016). The correlation between race and these socioeconomic and criminal history risk factors could result in higher risk scores and hence more intense levels of community corrections supervision and stricter penalties, including higher rates of pretrial detention for Black or Hispanic defendants (Hamilton, 2015; Harcourt, 2015; Klingele, 2016; Silver & Miller, 2002; Skeem & Lowenkamp, 2016; Starr, 2014; Stevenson & Mayson, 2017).

The issue of differential risk placement becomes particularly acute when it produces an imbalance of classification errors between White and non-White defendants (Chouldechova, 2016; Chouldechova & G'Sell, 2017; Kleinberg et al., 2016). By classification errors, we mean that the actuarial instrument could be classifying higher portions of nonrecidivating Black or Hispanic defendants into the high-risk categories (i.e., false positives) than their White counterparts. Conversely, higher rates of recidivating White defendants could be placed into the lower risk categories compared with similarly situated groups of recidivating Black or Hispanic defendants (i.e., false negatives; Chouldechova, 2016). This concern was recently underscored in the pretrial field by a study published by ProPublica examining the COMPAS algorithm for racial bias among pretrial defendants in Broward County, Florida. The study found nonrecidivating Black defendants being misclassified into the highest risk categories at rates 2 times higher than those of similarly situated non-Hispanic White defendants (Angwin, Larson, Mattu, & Kirchner, 2016; Chouldechova, 2016).

Although scholars have raised the alarm about errors in risk classification and race (see Harcourt, 2008; Starr, 2014, 2015), we focused on the issue of calibration rather than the balance rate of errors. Our decision to focus on calibration as opposed to error balance is informed by recent scholarship originating in the data science field on this issue (Berk, Heidari, Jabbari, Kearns, & Roth, 2017; Kleinberg et al., 2016). These studies examined whether it is possible to achieve the goals of both effective calibration and the balance of false positives and negatives by race, ethnicity, and sex. Ultimately, these studies found that it is not feasible to have an instrument that is both well calibrated and balanced except under certain highly constrained circumstances (Kleinberg et al., 2016). Specifically, either the recidivism base rates must be equal between the various race, ethnic, or sex subgroups or recidivism prediction needs to be perfect to meet the criteria of both effective calibration and equal error classification (Berk et al., 2017; Kleinberg et al., 2016). As these

requirements are almost never met, "the implications of this impossibility result are huge . . . [and] can lead to difficult stakeholder choices" (Berk et al., 2017, p. 19). In other words, attempting to rebalance error classifications will most likely result in a diminishment of risk prediction, while maximizing prediction will most likely produce inequality in error classifications. For these reasons, the current study assesses the PTRA's capacity to predict pretrial risk across race, Hispanic ethnicity, and sex groups. We direct those interested in the false positives, true positives, false negatives, and true negatives generated by the PTRA for defendant race, Hispanic ethnicity, and sex to the appendices in a longer version of this article available on the Social Science Research Network (SSRN).

Prediction rather than error classification is of paramount importance because the federal pretrial system requires a risk instrument capable of optimally assessing a defendant's likelihood of pretrial flight as well as danger to the community. If pretrial actuarial devices are going to be used by judicial actors, it is particularly important that they be free of predictive biases given the legal and moral controversies surrounding the decision to place defendants on pretrial detention. Hence, it is imperative to examine the PTRA's overall predictive capacities and most importantly for the issue of bias explore whether the instrument is well calibrated across race, Hispanic ethnicity, and sex subpopulations. Specifically, calibration entails that similar recidivism rates should manifest themselves for Whites and Blacks receiving the same risk score; for example, if high-risk Black and White defendants recidivate at the same rates, then the instrument is free from predictive bias, meaning it is well calibrated (Skeem & Lowenkamp, 2016). Conversely, predictive bias would be demonstrated if Blacks with the same risk scores as those of non-Hispanic Whites had lower likelihoods of engaging in criminal misconduct (Stevenson, in press). While there is not necessarily reason to believe the PTRA has predictive biases in terms of its capacity to predict pretrial violation outcomes equally well by defendant race, Hispanic ethnicity, and sex, there is also not necessarily reason to believe it does not.

Given the importance of the decisions made in pretrial contexts, it is incumbent upon the federal system to assess, as soon as reasonably possible, whether the PTRA's algorithm predicts pretrial violation outcomes equally well for Blacks and Hispanics compared with non-Hispanic Whites and females compared with males (Skeem & Lowenkamp, 2016; Skeem et al., 2016). These issues are of particular concern because historically many risk assessments were developed on nondiverse samples containing relatively few minorities or females, hence calling into question whether they can accurately predict reoffending behavior at the same level of accuracy for minorities and females as they do for defendants in the majority non-Hispanic White population (Skeem et al., 2016). Although the lack of diversity in assessment development has not been particularly important in the federal system, as sizable percentages of Blacks, Hispanics, and females were included in the PTRA construction and validation samples (see Lowenkamp & Whetzel, 2009), it is still important to monitor the performance of tools like the PTRA on these legally protected classes. A finding of predictive bias, for instance, could result in either Blacks or Hispanics being treated more harshly compared with non-Hispanic Whites by the pretrial system than is currently warranted. Moreover, a finding of sex bias could result in women being placed by the PTRA into higher risk categories than was warranted given their recidivistic behavior (Skeem & Lowenkamp, 2016; Skeem et al., 2016).

In general, those studies analyzing pretrial risk instruments under the calibration framework have failed to find evidence of predictive bias between non-Hispanic Whites and

Blacks although other research efforts have demonstrated that even well-calibrated risk instruments can still have biases in error classification and predictive parity. For examples of this emerging body of literature, see Chouldechova (2016) and Chouldechova and G'Sell (2017). Given our decision to focus on calibration, we highlight the empirical work on pretrial risk assessment and predictive bias while acknowledging that the issue of errors in risk classification represents an alternative means of evaluating risk tools for demographic biases. Northpointe, the company that owns COMPAS, for example, responded to ProPublica's study by demonstrating that their instrument has no predictive bias against Blacks compared with non-Hispanic White defendants (Dieterich, Mendoza, & Brennan, 2016; see also Flores, Bechtel, & Lowenkamp, 2016). In another example, analysis of the Virginia Pretrial Risk Assessment (VPRAI; see Danner, VanNostrand, & Spruance, 2016) found that this instrument predicted recidivism outcomes equally well for White and non-White defendants, which included any persons of color (i.e., Blacks, Hispanics, Asians, Native Americans, and other non-White races). While most studies of actuarial instruments have focused on the issue of predictive bias by race, some have explored whether these instruments produce biased predictions by sex. Of particular note, see Skeem et al. (2016) which, although focused on the postconviction stage, showed the federal Post Conviction Risk Assessment instrument overpredicting recidivism risk for female offenders.

PRESENT STUDY

The present study will endeavor to accomplish two primary goals. First, it will evaluate the PTRA's predictive efficacy by examining its capacity to predict pretrial violations involving rearrests for any or violent criminal activity, missed court appearances, or pretrial revocations among a national sample of federal defendants released pretrial. This part of the study is similar to and falls within existing research efforts attempting to validate a pretrial risk instrument for its overall predictive accuracy (Austin, Bhati, Jones, & Ocker, 2010; Blomberg, Bales, Mann, Meldrum, & Nedelec, 2010; Danner et al., 2016; Latessa, Lemke, Makarios, Smith, & Lowenkamp, 2010; Levin, 2010). Second, it will complement and augment the existing literature on the issue of risk assessment and bias by exploring whether the PTRA predicts pretrial violation outcomes equally well for non-Hispanic Whites, Blacks, and Hispanic defendants. Moreover, it will investigate the PTRA's predictive performance for male and female defendants. The results of the analysis investigating calibration will complement a small but growing body of literature that tests risk instruments' predictive biases across racial, ethnic, and sex subgroups (Danner et al., 2016; Dieterich et al., 2016; Skeem & Lowenkamp, 2016; Skeem et al., 2016).

Specifically, this article will endeavor to address the following research questions:

Research Question 1: How well does the PTRA perform predicting pretrial violation outcomes involving rearrests for any or violent criminal activity, revocations, or missed court appearances (e.g., FTA) for a national population of released federal defendants? What does the PTRA's predictive accuracy look like in terms of predicting any forms of pretrial violations or a combined new rearrest/FTA outcome?

Research Question 2: To what extent is the PTRA well calibrated by race, ethnicity, and sex (Skeem & Lowenkamp, 2016; Skeem et al., 2016)? Will the PTRA predict pretrial rearrest outcomes equally well for non-Hispanic Whites, Blacks, and Hispanic defendants? How well does the PTRA perform predicting pretrial rearrests for male and female defendants?

METHOD

PARTICIPANTS

Four samples were used to test the PTRA for its general predictive accuracy as well as examine the instrument for predictive bias. We first discuss the overall sample used to evaluate the instrument's general predictive capacities and then detail the matched sub-samples employed to examine the instrument for predictive bias between non-Hispanic Whites and Blacks, non-Hispanic Whites and Hispanics, and males and females.

Sample Used for Testing the PTRA's Overall Predictive Validity

The sample used to assess the PTRA's overall predictive validity was drawn from a larger population of 222,296 defendants who received PTRA assessments as part of their pretrial intake process between November 2009, when the PTRA was deployed in the federal system, and September 2015. This initial population included any defendants with PTRA assessments regardless of whether they were released or detained pretrial. Defendants were deemed eligible for this study if they (a) were released pretrial so that we could track their pretrial violation outcomes (*n* lost = 111,400 defendants), (b) no longer had a case in an opened status ensuring a complete measure of defendant violation activity while in the release phase (*n* lost = 24,376 defendants), and (c) had an actual PTRA assessment date for the purpose of tracking time while on pretrial release (*n* lost = 1,151 defendants). The use of these criteria yielded a pool of 85,369 defendants that could be used to evaluate the PTRA's predictive validity. It should be noted that while some pretrial risk assessment validation studies include opened or unresolved cases in their study sample, we opted to restrict our population to defendants with closed cases because our sample of over 85,000 defendants was more than sufficient to meet the study's objectives.

Table 1 provides a descriptive overview of defendants in the full PTRA validation sample. About two fifths of the study population (43%) comprised non-Hispanic Whites, while Blacks (26%) and Hispanics of any race (24%) accounted for similar portions of defendants. Males accounted for 72% of the study population, and the average defendant age was about 38 years (*SD* = 13.0). The majority of defendants in the study population (93%) were either U.S. born or naturalized citizens, a fact that should not be too surprising given that nearly all noncitizens are detained pretrial. Around 61% of defendants were classified into the lower PTRA risk categories (e.g., PTRA 1s and 2s), 25% were deemed moderate risk (PTRA 3s), and the remaining 14% were placed into the higher PTRA risk groups (e.g., PTRA 4s or 5s). Furthermore, the average PTRA score was 5.8 (*SD* = 2.5), with a range of 0 to 15 points. Last, defendants were on pretrial release for an average of 11 months (*SD* = 9.9).

Some caution is warranted in interpreting the information provided in Table 1 as it cannot be used to generalize to all defendants in the federal pretrial system. As stated previously, about half of defendants in the initial population could not be analyzed because they were detained pretrial and hence could not be tracked to assess their pretrial violation activity. In general, the population of detained defendants manifests higher PTRA risk characteristics compared with the released population. For instance, about 45% of detained defendants were classified into the PTRA 4 or 5 risk categories; in comparison, 14% of released defendants were designated into these higher PTRA risk groups. The divergence in risk scores

TABLE 1:   Descriptive Statistics of Federal Defendants in Study Sample

Variable	*n*	% or *M*
Race		
White, not Hispanic	35,581	42.8%
Black, not Hispanic	21,228	25.6
Hispanic, any race	20,112	24.2
Other race[a]	6,170	7.4
Sex		
Male	61,200	71.7%
Female	24,161	28.3
Citizenship		
U.S. citizen	73,601	86.8%
Naturalized U.S. citizen	4,802	5.7
Citizen of another country	6,406	7.6
PTRA risk categories		
1	28,033	32.8%
2	24,017	28.1
3	20,992	24.6
4	9,836	11.5
5	2,491	2.9
Age	85,356	37.8
		(13.0)
PTRA raw score	85,369	5.8
		(2.5)
Time on pretrial release (months)	85,335	11.3
		(9.9)
Number of defendants	85,369	

Note. Includes federal defendants released pretrial with PTRA assessments occurring between fiscal years 2010 and 2015. Standard deviations shown in parentheses. PTRA = pretrial risk assessment instrument.
[a]Other race includes Asians, Pacific Islanders, and Native Americans or Alaska Natives.

between the released and detained populations should not be too surprising, because defendants scoring on the higher end of the PTRA risk continuum are more likely to be detained pretrial compared with their lower risk counterparts. The selection of released defendants means that the study's findings are generalizable only to the population of released federal defendants and could possibly change if all defendants, encompassing both the released and detained populations, were included in the study population. The issue of measuring violation activity for detained defendants represents a perennial problem in pretrial research and is beyond the scope of this study. While acknowledging this methodological problem, assessing a risk instrument's predictive capacities for only released defendants represents a commonly applied and accepted approach to conducting pretrial risk assessment prediction and validation (see Austin et al., 2010; Cadigan et al., 2012; Danner et al., 2016; Latessa et al., 2010; Lowenkamp & Whetzel, 2009).

Generating Matched Subsamples for Testing the PTRA's Predictive Validity

From the larger sample of 85,369 defendants released pretrial, we created subsamples for testing the PTRA for predictive bias. The subsamples were generated by using a "one-to-one" matching procedure in which non-Hispanic Whites were matched with Blacks

and Hispanics on the criteria of sex and age, while males and females were matched on age and race/ethnicity. We decided to equalize the groups on age and sex because both factors have been shown to correlate with pretrial violations (Bechtel et al., 2011; Cohen & Reaves, 2007). Specifically, we used the ccmatch algorithm in STATA (see Cook, 2015) to generate matched race, ethnic, and sex subgroups for the bias analysis. This process resulted in matched samples of 41,112 non-Hispanic White (n = 20,556) and Black (n = 20,556) defendants, 30,622 non-Hispanic White (n = 15,311) and Hispanic (n = 15,311) defendants, and 40,888 male (n = 20,444) and female (n = 22,444) defendants. The process of selecting pairs of (matched) subsamples was conducted with replacement as participants were eligible for inclusion in repeated successive subsamples. For example, a White defendant could be used in both the matched White/Black subsample and the matched White/Hispanic subsample. In addition to matching on age, sex, and race/ethnicity, we investigated matching on the number of months released pretrial but decided against using this particular factor because it would have resulted in the loss of too many defendants. Rather, time on pretrial release was used as a regression covariate. We also excluded all noncitizens from the bias component of our analysis; the decision to exclude noncitizens is discussed subsequently.

Exclusion of Noncitizen Defendants

It is important to acknowledge that while we included the noncitizens for the analysis focusing on the PTRA's overall predictive capacities, we excluded this population from the bias component of this study. Our decision to remove the noncitizens is based on prior research showing that foreign-born individuals engage in criminal activity involving nonviolent or violent behavior less frequently than native-born individuals (Bersani & Piquero, 2016; Vaughn, Salas-Wright, DeLisi, & Maynard, 2014). While the lower rate of criminal offending would not be an issue if noncitizen defendants accounted for similar proportions of non-Hispanic Whites, Blacks, and Hispanics, that is not the case. Although similar percentages of non-Hispanic Whites (3%) and Blacks (3%)—$\chi^2(1)$ = 4.5; *ns*—in the initial sample were noncitizens, noncitizens comprised 18% of Hispanic defendants. Given that nearly a fifth of released Hispanics in the study population were noncitizens and considering the literature showing that nonnatives have lower criminal activity than their native-born counterparts (see Bersani & Piquero, 2016), a comparison of pretrial violation rates between Hispanics and other racial/ethnic categories without consideration of citizenship would be problematic. Hence, noncitizens were removed from the section of this article dealing with the issue of predictive bias. While the noncitizens were removed, it should be noted that naturalized citizens were kept in this analysis. Analyses were run with and without the naturalized citizens and produced no differences in the reported results.

MEASURES OF RISK

The PTRA's history, development, and risk-scoring scales have been discussed in other sections of this article and detailed in prior research (see Cadigan et al., 2012; Cadigan & Lowenkamp, 2011; Lowenkamp & Whetzel, 2009). To briefly reiterate, the PTRA was designed to predict pretrial violation outcomes involving rearrests for new criminal activity, FTA, and pretrial revocations. The instrument's algorithm assesses pretrial risk by having officers score defendants on their criminal history,

instant conviction offense, age, educational attainment, employment status, residential ownership, substance abuse problems, and citizenship status. Administration of the PTRA occurs prior to the initial hearing and during the intake process. The scores generated from the PTRA range from 0 to 15 and are used to place defendants into five different risk categories. For purposes of this study, we assess how the total PTRA scores and five categories perform in terms of risk prediction and calibration of risk across race, ethnic, and sex categories. We do not gauge this instrument's predictive capacities at the individual item or domain level.

MEASURING PRETRIAL VIOLATION OUTCOMES

For the section of this study focused on validating the PTRA's overall predictive efficacy, we examine whether this instrument effectively predicts rearrests for new offenses, rearrests for violent offenses, pretrial revocations, or FTAs. Pretrial revocations involve the removal of a defendant on pretrial release because of rearrests for new criminal activity or technical violations of release conditions, while FTAs imply the failure to show up to court for a designated hearing. Both violation outcomes were extracted from the Administrative Office of the U.S. Courts (AOUSC) internal case management database (Probation and Pretrial Services Automated Case Management System or PACTS).

Rearrests for new criminal activity, which we also refer to as pretrial recidivism,[2] were obtained from the National Crime Information Center (NCIC) and Access to Law Enforcement System (ATLAS). ATLAS is a software program used by the AOUSC that provides an interface for performing criminal record checks through a systematic search of official state and federal rap sheets (Baber, 2010). The ability to access and utilize official rap sheets represents a break from previous PTRA validation and other federal pretrial studies (see Cadigan et al., 2012; Cadigan & Lowenkamp, 2011; Cohen, 2013; Lowenkamp & Whetzel, 2009; VanNostrand & Keebler, 2009) where the pretrial rearrest data were inputted into the federal case management system by pretrial officers. This officer-inputted data did not provide as complete a picture of new criminal activity as that obtainable from official rap sheets. Pretrial rearrests are defined to include arrests for either felony or misdemeanor offenses (excluding arrests for technical violations) between the time of pretrial release and case closure. We also identified rearrests for violent offenses committed during the pretrial release phase. For violent rearrests, we used the definitions from the NCIC, which include homicide and related offenses, kidnapping, rape and sexual assault, robbery, and assault (Lowenkamp, Holsinger, & Cohen, 2015).

For the predictive bias component of this study, we focus our discussion on pretrial rearrests stemming from either any new criminal activity or violent offenses. Pretrial violations involving revocations or FTAs could be influenced by biased enforcement practices in the federal judicial system. For example, revocations for violations of technical pretrial conditions involve greater degrees of discretion among federal pretrial officers (Skeem & Lowenkamp, 2016; Skeem et al., 2016). Hence, a comparison of pretrial revocations between non-Hispanic Whites and Blacks could be diluted by the cultural norms, legal environment, or local policies and practices of judges and pretrial officers at the district court level. Conversely, rearrest activity for any or violent crimes is less open to subjective judicial and officer practices and hence provides a more objective criterion for examining the PTRA's predictive efficacy by race, ethnicity, and sex.

ANALYTICAL PLAN

To test for the PTRA's overall predictive capacities, we calculate descriptive statistics, effect sizes, and measures of predictive discrimination (e.g., areas under the receiver operating characteristic [ROC] curve [AUCs]). For the bias component of this study, we also used descriptive statistics (e.g., rearrest rates by PTRA risk category) and measures of discrimination (e.g., AUCs) to test for the PTRA's calibration across various demographic categories. Moreover, we employed multivariate logistic regressions with interaction terms to test whether the PTRA's predictive efficacy was moderated by defendant race, ethnicity, or sex (Skeem & Lowenkamp, 2016; Skeem et al., 2016). As these data encompassed a national sample of released defendants in 93 federal judicial districts (the District of Columbia maintains its own separate federal pretrial system), we clustered the standard errors at the district level to account for the nested nature of these data and the potential nonindependence of the standard errors (Hilbe, 2009). Given that even negligible differences can test at the standard .05 level because of the large sample sizes analyzed ($n = 30,000\text{-}40,000$ defendants depending upon matched sample sizes), we used a more conservative alpha level of .001 to denote statistical significance and reported effect sizes whenever possible. Last, it is important to note that there are various alternatives to assessing calibration which we do not use in this analysis. The expected (E)/observed (O) index (see Hanson, 2017), for example, has been suggested as another metric for analyzing an instrument's predictive capacity. This index measures the extent to which the anticipated number of recidivists matches the observed number, with a score of 1 indicating perfect calibration. Those interested in the E/O index metrics can reference the longer version of this article available on the SSRN.

RESULTS

EXAMINING THE PTRA'S OVERALL PREDICTIVE EFFECTIVENESS

Initially, we examine the PTRA's overall predictive efficacy for all released defendants in the sample ($N = 85,369$) regardless of race, ethnicity, or sex. Table 2 presents information on the percentage of released defendants committing pretrial violations involving revocations, new criminal rearrests, FTAs, or a combination of these outcomes across the five PTRA risk categories. The AUC-ROC scores are also presented as another measure of the PTRA's predictive accuracy. In the risk assessment literature, the AUC-ROC score provides an accepted gauge of an instrument's predictive accuracy in part because these scores, unlike correlations, are not influenced by low base rates (Babchishin & Helmus, 2016). This is especially important for the current study where the base rates for certain pretrial violation outcomes such as violent rearrests or FTAs are particularly low. Minimum AUC-ROC scores of 0.56, 0.64, and 0.71 correspond to "small," "medium," and "large" effects, respectively (Rice & Harris, 2005).

Results from Table 2 show that the PTRA effectively predicts pretrial violations irrespective of whether the outcome of interest involves the revocation from pretrial release, the rearrest for any felony or misdemeanor offenses, the rearrest for violent offenses, the FTA in court, or the combination of these outcomes. For example, the percentage of defendants with any adverse events—meaning they had a revocation, new rearrest, or FTA—while on pretrial release increased in the following incremental fashion by PTRA risk category: 5% (PTRA 1s), 11% (PTRA 2s), 20% (PTRA 3s), 29% (PTRA 4s), and 36% (PTRA 5s). These results were in the anticipated direction of higher failure rates for each increase in risk

TABLE 2:   PTRA Failure Rates for Any Adverse Events Involving New Criminal Arrests, Pretrial Revocations, or Failure to Appear

PTRA risk categories	n	Any adverse event (%)	Pretrial revocation (%)	New arrest or FTA (%)	Arrests for any offense (%)	Arrests for violent offenses (%)	FTA (%)
1	28,033	4.7	2.0	3.2	2.6	0.3	0.7
2	24,017	10.5	5.5	6.4	5.1	0.7	1.5
3	20,992	19.9	12.6	10.6	8.5	1.3	2.5
4	9,836	29.1	18.6	15.6	13.0	2.1	3.2
5	2,491	36.1	22.5	19.8	16.5	2.9	4.6
Base rate	85,369	13.8	8.1	7.8	6.4	1.0	1.7
AUC, PTRA total		0.71	0.73	0.68	0.68	0.69	0.67
99.9% CI		[0.71, 0.72]	[0.72, 0.74]	[0.67, 0.69]	[0.66, 0.69]	[0.66, 0.72]	[0.65, 0.69]

Note. Any adverse event includes pretrial violations involving a new criminal arrest, failure to make court appearances, or pretrial revocations. Specific failure events will not sum to totals as defendants can experience multiple violation types simultaneously. PTRA = pretrial risk assessment instrument; AUC = area under the receiver operating characteristic curve; FTA = failure to appear; CI = confidence interval.

classification. Moreover, the PTRA risk scale manifested an AUC-ROC score of 0.71 (99.9% confidence interval [CI] = [0.71, 0.72]) for the any adverse event outcome, meaning that this instrument provides good to excellent capacities in terms of predicting all forms of pretrial violations (Desmarais & Singh, 2013; Rice & Harris, 2005).

Similar patterns were revealed about the PTRA's capacities for predicting specific forms of pretrial violations, including rearrests for any or violent offenses, FTAs, or pretrial revocations. For instance, the percentage of defendants rearrested for any offenses while on pretrial release was 3% for PTRA 1s, 5% for PTRA 2s, 9% for PTRA 3s, 13% for PTRA 4s, and 17% for PTRA 5s. In addition to examining failure rates by risk category, an overview of the AUC-ROC scores shows them ranging from 0.67 to 0.73 for the FTA (0.67, 99.9% CI = [0.65, 0.69]), any rearrests (0.68, 99.9% CI = [0.66, 0.69]), violent rearrests (0.69, 99.9% CI = [0.66, 0.72]), combined rearrest/FTA (0.68, 99.9% CI = [0.67, 0.69]), or pretrial revocations (0.73, 99.9% CI = [0.72, 0.74]) outcomes. These scores mean that the PTRA provides "good" to "excellent" predictive capacities for these specific types of pretrial violations (Desmarais & Singh, 2013).

TESTING FOR PREDICTIVE BIAS BETWEEN NON-HISPANIC WHITES AND BLACKS

While previously we have demonstrated the PTRA's predictive efficacy for all released defendants, in this and subsequent sections, we focus on the issue of predictive bias. Specifically, we hypothesized that the PTRA will have the same predictive meaning in terms of a defendant's likelihood of committing pretrial recidivism regardless of their race, ethnicity, or sex. In this section, we focus on whether the PTRA predicts pretrial rearrests (any or violent) equally well for non-Hispanic Whites and Blacks. Because rearrests, particularly for violent behavior, are considered more objective outcome measures compared with other forms of pretrial violations (see Piquero & Brame, 2008; Skeem & Lowenkamp, 2016; Skeem et al., 2016), these findings focus on the relationship between pretrial rearrests, race, and risk. In general, our analysis supports the hypothesis that the PTRA strongly predicts pretrial rearrests for both non-Hispanic

TABLE 3:   PTRA Rearrest Rates (Any or Violent) Between Non-Hispanic Whites and Blacks

PTRA risk categories	Arrests for any offense				Arrests for violent offenses			
	White		Black		White		Black	
	n	% arrested	*n*	% arrested	*n*	% arrested	*n*	% arrested
1	7,271	3.1	5,001	2.8	7,271	0.3	5,001	0.4
2	5,797	5.7	5,703	6.7	5,797	0.7	5,703	1.2
3	4,727	9.8	5,489	10.7	4,727	1.5	5,489	1.9
4	2,257	14.8	3,272	15.7	2,257	1.7	3,272	3.5
5	504	15.1	1,091	19.4	504	2.6	1,091	3.8
Base rate	20,556	6.9	20,556	8.9	20,556	0.9	20,556	1.7
AUC, PTRA total	0.68		0.67		0.69		0.69	
99.9% CI	[0.65, 0.70]		[0.65, 0.70]		[0.64, 0.75]		[0.64, 0.73]	

Note. Includes 41,112 non-Hispanic White (*n* = 20,556) and Black (*n* = 20,556) defendants on pretrial release. Non-Hispanic Whites and Blacks matched on age and sex. PTRA = pretrial risk assessment instrument; AUC = area under the receiver operating characteristic curve; CI = confidence interval.

Whites and Blacks and that PTRA scores are associated with similar probabilities of rearrest regardless of race.

Strength of Prediction

In Table 3, we examine whether the association between the PTRA risk scores and pretrial rearrest varied between non-Hispanic White and Black defendants. Results show similar patterns of rearrest activity involving any or violent offenses for both race groups. Specifically, the any and violent rearrest outcomes manifested monotonically increasing rates of rearrest irrespective of the defendant's racial background. It is interesting to note that the rearrest rates involving any offenses were similar for non-Hispanic Whites with PTRA 4 and 5 risk scores; Blacks, however, manifested increasing rearrest rates between these high-risk categories.

Table 3 also presents the AUC-ROC values generated for the total PTRA scores by race. The AUC-ROC scores are particularly informative of risk prediction given the low base rates of violent offending activity for released federal defendants. An examination of the AUC-ROC scores evidences that the instrument has "good" or "moderate" predictive capacities across defendant race as these scores ranged from 0.67 (99.9% CI = [0.65, 0.70]) to 0.69 (99.9% CI = [0.64, 0.75]) among the any and violent rearrest outcomes. The differences in the AUC-ROC scores between Blacks and non-Hispanic Whites for violent rearrests—$\chi^2(1) = 0.11$; *ns*—or any rearrests, $\chi^2(1) = 0.0$; *ns*—were not statistically significant.

Form of Prediction

As we have shown that the PTRA predicts pretrial rearrests for any and violent criminal activity equally well among Black and non-Hispanic White defendants, we subsequently examined whether the form of the relationship between the PTRA risk scores and pretrial recidivism differs by race. Stated another way, this analysis investigates whether non-Hispanic Whites and Blacks manifest similar rearrest likelihoods for any

TABLE 4: Logistic Regression Models Testing the Predictive Fairness of PTRA Between Non-Hispanic Whites and Blacks

Model variables	Model 1 Odds ratio	Model 1 99.9% CI Lower	Model 1 99.9% CI Upper	Model 2 Odds ratio	Model 2 99.9% CI Lower	Model 2 99.9% CI Upper	Model 3 Odds ratio	Model 3 99.9% CI Lower	Model 3 99.9% CI Upper	Model 4 Odds ratio	Model 4 99.9% CI Lower	Model 4 99.9% CI Upper
Any pretrial rearrest												
Black	1.32*	1.14	1.52	—	—	—	1.07	0.94	1.23	1.15	0.76	1.72
PTRA risk score	—	—	—	1.30*	1.27	1.34	1.30*	1.26	1.34	1.31*	1.25	1.36
PTRA Score × Black	—	—	—	—	—	—	—	—	—	0.99	0.94	1.05
Time on release	1.02*	1.02	1.03	1.03*	1.02	1.04	1.03*	1.02	1.04	1.03*	1.02	1.04
(Constant)	0.05			0.01			0.01			0.01		
Model χ²	128.7			937.4			932.7			933.1		
Log-likelihood	−11,244.9			−10,650.4			−10,648.7			−10,648.5		
Pseudo R²	0.01			0.06			0.06			0.06		
Violent pretrial rearrest												
Black	1.89*	1.40	2.55	—	—	—	1.51*	1.11	2.06	1.74	0.68	4.47
PTRA risk score	—	—	—	1.33*	1.27	1.40	1.32*	1.25	1.39	1.33*	1.23	1.45
PTRA Score × Black	—	—	—	—	—	—	—	—	—	0.98	0.87	1.10
Time on release	1.02*	1.01	1.04	1.03*	1.02	1.04	1.03*	1.02	1.04	1.03*	1.02	1.04
(Constant)	0.01			0.00			0.00			0.00		
Model χ²	102.58			354.2			383.0			387.7		
Log-likelihood	−2,777.5			−2,669.3			−2,659.2			−2,659.1		
Pseudo R²	.02			.05			.06			.06		

Note. Models include 41,100 non-Hispanic White and Black defendants on pretrial release. Non-Hispanic Whites and Blacks matched on age and sex. PTRA = pretrial risk assessment instrument; CI = confidence interval; — = not applicable.
*$p < .001$.

given PTRA score net of statistical controls. This section also explores whether the shape (i.e., the regression slopes) between the PTRA risk scores and rearrest odds for any or violent offenses is moderated by the defendant's race (Skeem & Lowenkamp, 2016; Skeem et al., 2016). Preferably, the form of the relationship between the PTRA risk scores and defendant rearrest odds will be similar between non-Hispanic Whites and Blacks; moreover, each race group should manifest similar rearrest probabilities for any given PTRA risk score. It should be noted that all models include time on pretrial release as a statistical control, and these models were run using a procedure in which robust standard errors were clustered by district (Hilbe, 2009).

A series of logistic regression models were employed to examine these issues (four models for any pretrial rearrests and four for violent pretrial rearrests). The models were employed to test for differences in the regression slopes and intercepts across the two race categories (Skeem & Lowenkamp, 2016; Skeem et al., 2016). As shown in Table 4, Models 1 and 2 include only the defendant race (Model 1) or the PTRA score (Model 2), Model 3 includes both the PTRA score and race, and Model 4 incorporates race, PTRA score, and an interaction term of defendant race and PTRA score.

Results from the models are displayed in Table 4. Regarding slope, we show that the shape or form of the relationship between the PTRA scores and pretrial recidivism was the same for both non-Hispanic Whites and Blacks. This finding is supported by the fact that the interaction terms do not significantly improve either the any—$\Delta\chi^2(1) = 0.30$; *ns*—or violent pretrial rearrest models, $\Delta\chi^2(1) = 0.29$; *ns*. Moreover, the odds ratios for
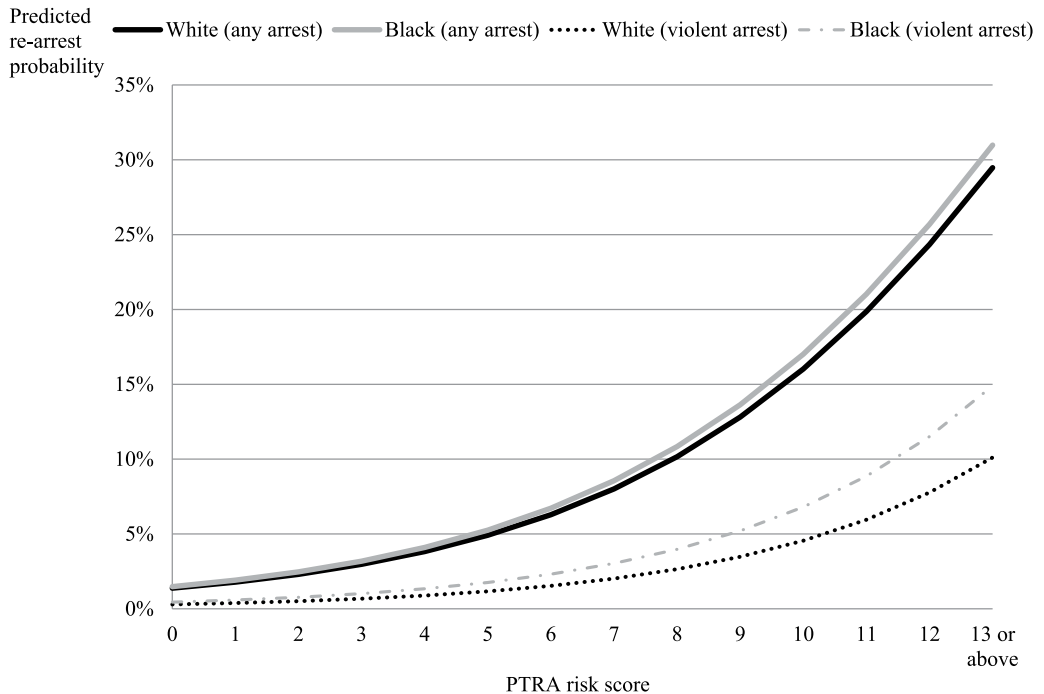
Figure 1:   Predicted Probabilities of New Pretrial Rearrests for Any or Violent Offenses by PTRA Score Between Non-Hispanic Whites and Blacks
Note. PTRA scores above 13 were recoded into scores of 13 as there were not enough defendants across the race/ethnic categories to produce statistically reliable estimates. Predicted violent rearrest probabilities weighted by a factor of 2 because of low base rates. PTRA = pretrial risk assessment instrument.

the interaction terms in either rearrest model evidence essentially trivial effects regarding the potential for race to moderate the relationship between PTRA scores and recidivism.

In terms of intercept differences, we find that Blacks and non-Hispanic Whites had statistically similar intercepts for the models examining the association between PTRA risk scores and any pretrial recidivism; however, the intercept of the relationship between PTRA risk scores and violent pretrial recidivism was significantly lower for non-Hispanic Whites compared with Black defendants. These findings are supported by the fact that race did not add any predictive utility to the any pretrial rearrest model, $\Delta\chi^2(1) = 3.1$; *ns*—but did statistically enhance the violent pretrial rearrest model—$\Delta\chi^2(1) = 19.1$; $p < .001$.

Another way of illustrating the form of the association between PTRA risk scores and pretrial recidivism involves calculating the average predicted probabilities of pretrial rearrest (any or violent) for non-Hispanic Whites and Blacks by the individual PTRA scores (see Figure 1). The predicted probabilities for the any or violent pretrial rearrest outcomes were calculated based on the Model 3 logistic regressions. Given the very low base rates

TABLE 5:   PTRA Rearrest Rates Between Non-Hispanic Whites and Hispanics

| | Arrests for any offense | | | | Arrests for violent offenses | | | |
| | White | | Hispanic | | White | | Hispanic | |
PTRA risk categories	n	% arrested	n	% arrested	n	% arrested	n	% arrested
1	4,836	2.9	3,367	2.2	4,836	0.3	3,367	0.2
2	4,453	5.8	4,690	4.1	4,453	0.6	4,690	0.5
3	3,775	10.3	4,742	6.8	3,775	1.5	4,742	0.9
4	1,824	14.6	2,084	10.0	1,824	1.7	2,084	1.2
5	423	13.2	428	14.5	423	2.1	428	3.0
Base rate	15,311	7.3	15,311	5.6	15,311	0.9	15,311	0.7
AUC, PTRA total	0.67		0.65		0.67		0.66	
99.9% CI	[0.65, 0.70]		[0.62, 0.68]		[0.60, 0.74]		[0.58, 0.74]	

Note. Includes 30,622 non-Hispanic White (n = 15,311) and Hispanic (n = 15,311) defendants on pretrial release. Whites and Hispanics matched on age and sex. Number and pretrial violation rates for White defendants will not match those in Table 3 as the matched samples for White defendants differ between the two tables. PTRA = pretrial risk assessment instrument; AUC = area under the receiver operating characteristic curve; CI = confidence interval.

associated with pretrial violent rearrests, we weighted the violent rearrest predicted probabilities by a factor of 2. According to Figure 1, the probabilities of pretrial rearrest (any or violent) behave similarly for both non-Hispanic Whites and Blacks; each group of defendants witnesses curvilinear increases in their rearrest probabilities by PTRA risk score. Conversely, the intercepts are the same between non-Hispanic Whites and Blacks for the any rearrest predicted probabilities, but are lower for non-Hispanic Whites than Blacks among the violent rearrest predicted probabilities.

TESTING FOR PREDICTIVE BIAS BETWEEN NON-HISPANIC WHITES AND HISPANICS (ANY RACE)

This part of the study focuses on whether the PTRA predicts pretrial rearrests (any or violent) equally well for non-Hispanic Whites and Hispanics. While we show that the PTRA strongly predicts pretrial rearrests for both groups, the instrument overpredicts the odds of any but not violent pretrial recidivism among Hispanic defendants. In other words, the PTRA overestimates the likelihood of Hispanic defendants being rearrested for any offenses; however, the instrument produces estimations on the odds of violent rearrest that are essentially the same by defendant ethnicity.

Strength of Prediction

In Table 5, we examine whether the association between the PTRA risk scores and pretrial rearrest varied between non-Hispanic White and Hispanic defendants. Results show non-Hispanic Whites and Hispanics having similar patterns of any or violent rearrest activity. Specifically, the any and violent rearrest rates increased in a systematic fashion across the PTRA risk categories among both sets of defendants. Table 5 also presents the AUC-ROC values. The AUC-ROC scores show that the instrument's predictive capacities fell into the "good" or "moderate" range; these values ranged from 0.65 (99.9% CI = [0.62, 0.68]) to 0.67 (99.9% CI = [0.60, 0.74]) depending upon the rearrest outcome examined. The differences in the AUC-ROC scores between Hispanics and non-Hispanic Whites for

TABLE 6: Logistic Regression Models Testing the Predictive Fairness of PTRA Between Non-Hispanic Whites and Hispanics

Model variables	Model 1 Odds ratio	99.9% CI Lower	99.9% CI Upper	Model 2 Odds ratio	99.9% CI Lower	99.9% CI Upper	Model 3 Odds ratio	99.9% CI Lower	99.9% CI Upper	Model 4 Odds ratio	99.9% CI Lower	99.9% CI Upper
Any pretrial rearrest												
Hispanic	0.77	0.55	1.09	—	—	—	0.69*	0.51	0.94	0.72	0.36	1.41
PTRA risk score	—	—	—	1.29*	1.24	1.34	1.29*	1.25	1.34	1.30*	1.24	1.35
PTRA Score × Hispanic	—	—	—	—	—	—	—	—	—	1.00	0.92	1.07
Time on release	1.02*	1.01	1.03	1.03*	1.02	1.04	1.03*	1.02	1.04	1.03*	1.02	1.04
(Constant)	0.06			0.01			0.01			0.01		
Model χ²	76.1			562.5			658.6			673.8		
Log-likelihood	−7,233.2			−6,939.9			−6,910.5			−6,910.5		
Pseudo R²	.01			.05			.05			.05		
Violent pretrial rearrest												
Hispanic	0.78	0.47	1.30	—	—	—	0.71	0.46	1.09	0.59	0.16	2.12
PTRA risk score	—	—	—	1.30*	1.19	1.42	1.31*	1.20	1.43	1.29*	1.17	1.43
PTRA Score × Hispanic	—	—	—	—	—	—	—	—	—	1.03	0.87	1.21
Time on release	1.02*	1.01	1.04	1.03*	1.01	1.04	1.03*	1.02	1.04	1.03*	1.02	1.04
(Constant)	0.01			0.00			0.00			0.00		
Model χ²	39.6			142.7			189.1			188.2		
Log-likelihood	−1,440.2			−1,396.1			−1,392.4			−1,392.3		
Pseudo R²	0.01			0.04			0.04			0.04		

Note. Models include 30,610 non-Hispanic White and Hispanic defendants on pretrial release. Non-Hispanic Whites and Hispanics matched on age and sex. PTRA = pretrial risk assessment instrument; CI = confidence interval; — = not applicable.
*$p < .001$.

violent rearrests—$\chi^2(1) = 0.13$; *ns*—or any rearrests—$\chi^2(1) = 2.62$; *ns*—were not statistically significant.

Form of Prediction

Next, we investigate whether the form of the relationship between the PTRA risk scores and pretrial recidivism differs by ethnicity. Basically, we examined whether non-Hispanic Whites and Hispanics had similar rearrest likelihoods for any given PTRA score (i.e., regression intercepts) and explored whether defendant ethnicity moderated the shape of the relationship (i.e., regression slopes) between the PTRA scores and rearrest odds (Skeem & Lowenkamp, 2016; Skeem et al., 2016).

Several logistic regression models were employed to examine the relationship between Hispanic ethnicity, risk, and recidivism (four models for any pretrial rearrests and four for violent pretrial rearrests; Table 6). These analyses show that the form of the relationship between the PTRA scores and pretrial recidivism is the same for both non-Hispanic Whites and Hispanics, a finding supported by the fact that neither interaction terms significantly contributed to the any—$\Delta\chi^2(1) = 0.03$; *ns*—or violent pretrial rearrest models—$\Delta\chi^2(1) = 0.24$; *ns*. In terms of intercept differences, we find that Hispanics and non-Hispanic Whites had statistically similar intercepts for the models examining the association between PTRA risk scores and violent pretrial recidivism; however, the intercept of the relationship between PTRA risk scores and any pretrial recidivism was significantly lower for Hispanics compared with non-Hispanic White defendants. Support for these findings is demonstrated by
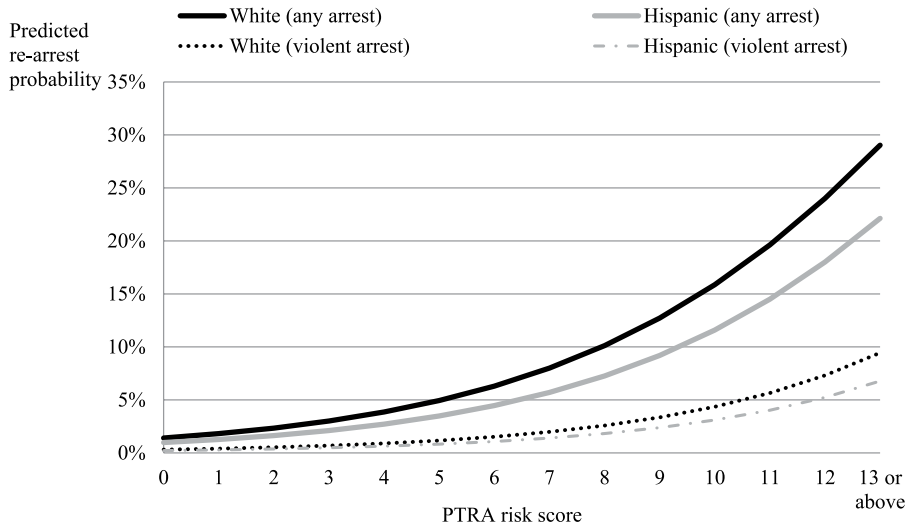
Figure 2:   Predicted Probabilities of New Pretrial Rearrests for Any or Violent Offenses by PTRA Score Between Non-Hispanic Whites and Hispanics
Note. PTRA scores above 13 were recoded into scores of 13 as there were not enough defendants across the race/ethnic categories to produce statistically reliable estimates. Predicted violent rearrest probabilities weighted by a factor of 2 because of low base rates. PTRA = pretrial risk assessment instrument.

the fact that Hispanic ethnicity did not increase the predictive utility for the violent pretrial rearrest model—$\Delta\chi^2(1) = 6.9$; *ns*—but did statistically enhance the any pretrial rearrest model—$\Delta\chi^2(1) = 15.9$; $p < .001$.

Figure 2 displays the form of the association between PTRA risk scores and pretrial recidivism involving any or violent offenses for non-Hispanic Whites and Hispanics by the individual PTRA scores. The predicted probabilities for the any or violent pretrial rearrest outcomes were calculated based on the Model 3 logistic regressions. As relatively few defendants were rearrested for violent offenses, we weighted the violent rearrest predicted probabilities by a factor of 2. Basically, the probabilities of pretrial rearrest (any or violent) manifest similar patterns of change for each one-point increase in the PTRA score. Curvilinear relationships between the PTRA risk score and rearrest probability were observed among both non-Hispanic White and Hispanic defendants. Conversely, non-Hispanic Whites and Hispanics had similar rearrest probabilities for offenses involving violent conduct, but the probability of rearrest for any offense was lower for Hispanics compared with non-Hispanic defendants.

TESTING FOR PREDICTIVE BIAS BETWEEN MALES AND FEMALES

Last, we examine the PTRA's capacity to predict pretrial rearrests (any or violent) for male and female defendants. We investigated the form or shape of the relationship between risk scores and recidivism and analyzed whether the intercepts differed between male and female defendants. In findings mirroring that of other analyses, we show that the PTRA

TABLE 7: PTRA Rearrest Rates Between Males and Females

PTRA risk categories	Arrests for any offense				Arrests for violent offenses			
	Male		Female		Male		Female	
	n	% arrested	n	% arrested	n	% arrested	n	% arrested
1	6,189	3.0	6,872	2.0	6,189	0.3	6,872	0.2
2	5,494	5.7	6,011	5.2	5,494	1.0	6,011	0.6
3	5,186	10.1	5,124	7.9	5,186	1.7	5,124	0.7
4	2,813	13.9	2,019	13.0	2,813	2.4	2,019	1.3
5	762	18.2	418	15.3	762	3.8	418	1.0
Base rate	20,444	7.6	20,444	5.8	20,444	1.3	20,444	0.6
AUC, PTRA total	0.67		0.69		0.69		0.66	
99.9% CI	[0.65, 0.70]		[0.66, 0.71]		[0.64, 0.74]		[0.59, 0.74]	

Note. Includes 40,888 male (n = 20,444) and female (n = 22,444) defendants on pretrial release. Males and females matched on age and race/ethnicity (White, Black, Hispanic). Defendants in the other race categories excluded. PTRA = pretrial risk assessment instrument; AUC = area under the receiver operating characteristic curve; CI = confidence interval.

predicts pretrial recidivism activity for both males and females; however, the instrument does overpredict pretrial rearrests for violent offenses among female defendants. There were no differences in the instrument's capacity to predict pretrial rearrests for any offenses between males and females.

Strength of Prediction

Results examining the association between PTRA risk scores and pretrial rearrest (any or violent) for male and female defendants are reported in Table 7. These findings generally show the pretrial rearrest rates for any or violent offenses increasing in a monotonic fashion by PTRA risk level among both sets of defendants. An exception to this pattern involves the female violent rearrest rates which were similar for females with PTRA 4 and 5 scores, although the female violent rearrest rates do systemically increase between PTRA Categories 1 through 4. Table 7 also displays the AUC-ROC values. The AUC-ROC scores ranged from 0.66 (99.9% CI = [0.59, 0.74]) to 0.69 (99.9% CI = [0.66, 0.71]) for the any and violent rearrest outcomes, meaning that the instrument has "good" or "moderate" predictive capacities for both sexes. Males and females manifested similar AUC-ROC values for the violent rearrests—$\chi^2(1)$ = 0.8; *ns*) and any rearrests—$\chi^2(1)$ = 1.5; *ns*—outcomes.

Form of Prediction

Similar to the race and ethnicity analysis, we investigated whether males and females had comparable rearrest likelihoods for any given PTRA score (i.e., regression intercepts) and analyzed whether sex moderated the shape of the relationship (i.e., regression slopes) between the PTRA scores and rearrest odds (Skeem & Lowenkamp, 2016; Skeem et al., 2016). A finding showing neither interaction terms contributed significantly to the any—$\Delta\chi^2(1)$ = 3.9; *ns*—or violent pretrial rearrest models—$\Delta\chi^2(1)$ = 0.4; *ns*—demonstrates that the form of the relationship between the PTRA scores and pretrial recidivism was the same for both males and females. As for differences in intercepts, we show males and females

TABLE 8: Logistic Regression Models Testing the Predictive Fairness of PTRA Between Males and Females

PTRA risk categories	Model 1 Odds ratio	Model 1 99.9% CI Lower	Model 1 99.9% CI Upper	Model 2 Odds ratio	Model 2 99.9% CI Lower	Model 2 99.9% CI Upper	Model 3 Odds ratio	Model 3 99.9% CI Lower	Model 3 99.9% CI Upper	Model 4 Odds ratio	Model 4 99.9% CI Lower	Model 4 99.9% CI Upper
Any pretrial rearrest												
Male	1.29*	1.12	1.50	—	—	—	1.15	0.99	1.34	1.48	0.95	2.30
PTRA risk score	—	—	—	1.32*	1.28	1.36	1.31*	1.27	1.36	1.34*	1.27	1.41
PTRA Score × Male	—	—	—	—	—	—	—	—	—	0.97	0.91	1.02
Time on release	1.02*	1.02	1.03	1.03*	1.02	1.04	1.03*	1.02	1.04	1.03*	1.02	1.04
(Constant)	0.05			0.01			0.01			0.01		
Model χ^2	133.5			867.7			986.8			975.9		
Log-likelihood	−9,914.3			−9,378.3			−9,372.1			−9,370.1		
Pseudo R^2	.01			.07			.07			.07		
Violent pretrial rearrest												
Male	2.23*	1.55	3.21	—	—	—	1.98*	1.37	2.85	1.61	0.54	4.82
PTRA risk score	—	—	—	1.33*	1.24	1.43	1.31*	1.22	1.40	1.28*	1.13	1.46
PTRA Score × Male	—	—	—	—	—	—	—	—	—	1.03	0.89	1.19
Time on release	1.02*	1.01	1.03	1.03*	1.02	1.04	1.03*	1.02	1.04	1.03*	1.02	1.04
(Constant)	0.00			0.00			0.00			0.00		
Model χ^2	121.3			188.9			257.8			268.9		
Log-likelihood	−2,072.8			−2,011.7			−1,992.7			−1,992.5		
Pseudo R^2	.0			.05			.06			.06		

Note. Models include 40,873 male and female defendants on pretrial release. Males and females matched on age and race/ethnicity (White, Black, Hispanic). PTRA = pretrial risk assessment instrument; CI = confidence interval; — = not applicable.
*$p < .001$.

having statistically similar intercepts for the models examining the association between PTRA risk scores and any pretrial rearrests—$\Delta\chi^2(1) = 9.7$; *ns*—but the intercept of the relationship between PTRA risk scores and violent pretrial rearrests was significantly lower for females compared with males, $\Delta\chi^2(1) = 37.2$; $p < .001$ (Table 8).

Figure 3 displays the form of the association between PTRA risk scores and pretrial recidivism involving any or violent offenses for males and females by the individual PTRA scores. As in the prior figures, these predicted probabilities were calculated based on the Model 3 logistic regressions and the violent offense predicted probabilities were weighted by a factor of 2. These predicted probabilities showed the form or slope of the relationship between the PTRA risk scores and rearrests was similar for both males and females regardless of whether the criminal conduct involved any or violent offenses. Moreover, males and females had similar rearrest probabilities for any offenses, but females were less likely to be rearrested for violent offenses than males.

DISCUSSION

In a rather clear and direct manner, the current research indicates that the PTRA is a valid predictor of several important pretrial outcomes. For each of the six outcomes tested, the failure rates increase, monotonically, when moving from one risk category to the next. Furthermore, the failure rates associated with each category for each of the six outcomes are, for the most part, practically meaningful. The two outcomes that might be exceptions to this are rearrest for violent offenses and FTA as the overall base rate for these
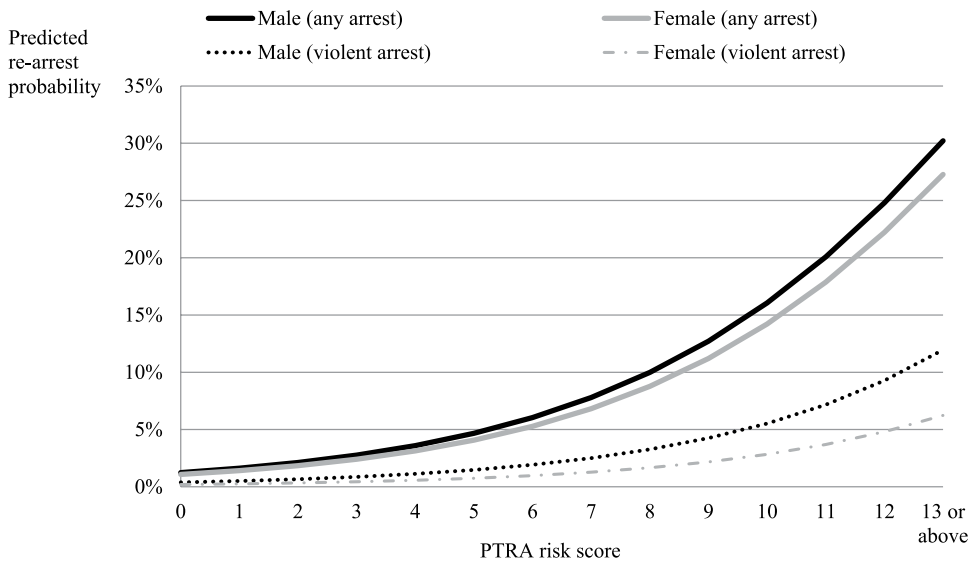
Predicted
re-arrest
probability

Figure 3:   Predicted Probabilities of New Pretrial Rearrests for Any or Violent Offenses by PTRA Score Between Males and Females
Note. PTRA scores above 13 were recoded into scores of 13 as there were not enough defendants across the sex categories to produce statistically reliable estimates. Predicted violent rearrest probabilities weighted by a factor of 2 because of low base rates. PTRA = pretrial risk assessment instrument.

two outcomes is 1.0% and 1.7% respectively. Even so, for each of the two aforementioned outcomes, Category 1 defendants are associated with a failure rate that is roughly one third the overall base rate, and Category 5 defendants are associated with a failure rate that is roughly 3 times the base rate. While it might be difficult to make practical use of the differences in failure rates between two contiguous categories, when spanning the full range of the scale, the categorizations are meaningful. In addition to the practical assessment of usefulness (i.e., meaningful differences in failure rates across risk categories), we calculated the AUC-ROC score to assess the PTRA's capacity to discriminate between recidivists and nonrecidivists. The AUC-ROC values between the PTRA total and the various outcomes were all in the "good" to "excellent" range (or "medium" to "large" depending on which nomenclature for AUC-ROC values is used).

In summary, when considering the PTRA for all defendants, it appears that the PTRA is a valid predictor of any adverse event, pretrial revocation, new rearrest, FTA, and rearrest for a violent offense. It is remarkable and worth noting that one score can predict this variety of outcomes. Recent developments in pretrial risk assessment have shifted toward the development of specific scales that maximize the prediction of different outcomes (LJAF, 2016). However, it might be that the simplicity of a single score, the relative accuracy in predicting various outcomes with a single score, and the limitations of data available for scale construction and administration make single-score assessments a continued viable option.

Analyses assessing the function of the PTRA across matched samples of non-Hispanic White and Black defendants indicate that the PTRA operates similarly for these two groups

of defendants. While there are variations in the failure rates across groups by race at any given risk category, these differences are, for the most part, slight and practically negligible. Basically, as the PTRA categories increase, so do the rearrest rates for both non-Hispanic Whites and Blacks, with the rearrest rates leveling off for Whites but not Blacks at the PTRA 4 category. For both groups of defendants, the PTRA produces AUC-ROC values that are nearly identical and are in the "good" range. Furthermore, interaction terms between the PTRA score and race from logistic regression models predicting rearrest for any reason and rearrest for a violent offense were not statistically significant. It is also worth noting that in the fully specified models race was not a significant predictor in terms of predicting any rearrests, although it was a significant predictor, with Blacks more likely to be rearrested than non-Hispanic Whites, in the prediction of violent rearrests. These findings are mostly consistent with research on the postconviction risk assessment in use in the U.S. probation system (see Skeem & Lowenkamp, 2016) and research being generated with other risk assessments (see Brennan, Dieterich, & Ehret, 2009; Dieterich et al., 2016; Flores et al., 2016; Lowenkamp & Bechtel, 2007).

Turning to the results assessing the prediction of the PTRA across Hispanic origin, the trends are generally similar. In particular, both groups manifest similar rearrest behavior by PTRA risk category, with rearrest rates increasing by risk category until the PTRA 4 classification is reached; afterward, rearrest rates continue to rise for Hispanics, but not non-Hispanic Whites. The AUC-ROC values, based on 99.9% CIs, are of the same for both outcomes (rearrest for any reason and rearrest for a violent offense) and continue to fall in the "good" or "medium" range. Logistic regression models indicate that the PTRA performs similarly for Hispanic and non-Hispanic White defendants and further that the interaction term between Hispanic origin and the total PTRA score is not significant. Notwithstanding these findings, it appears that it might be the case that the matched sample of Hispanic defendants has lower failure rates across the two outcomes. This difference is more or less pronounced depending on which category of the PTRA is being compared. This difference is also evinced by the odds ratio from Model 3 (predicting any rearrest), which indicates that on average Hispanic defendants have a lower likelihood of being rearrested for any offenses than do their matched non-Hispanic White counterparts. The odds of garnering a violent arrest, however, were statistically similar between non-Hispanic Whites and Hispanics. While a lack of consistent and statistically significant differences between Hispanic and non-Hispanic White defendants provides evidence that the PTRA is equally valid among both groups of defendants, future research might monitor differences in failure rates between these two groups of defendants.

Finally, analyses comparing the performance of the PTRA between males and females were completed. A matched sample of males and females produced AUC-ROC values that were similar and achieved magnitudes in the same range ("good" or "medium") as observed in the total sample and the other matched groups. Logistic regression results did not yield significant interaction terms, meaning the assessment performs similarly for males and females. This is not too surprising given the extant research on this topic (see, for example, Monahan, Skeem, & Lowenkamp, 2017; Smith, Cullen, & Latessa, 2009). What is somewhat surprising is that the odds ratios for being male were not consistently statistically significant. Specifically, the odd ratios indicated similar rearrest probabilities for males and females regarding any offenses, but males had significantly higher likelihoods of being arrested for violent offenses than females. We expected to see males uniformly failing at

different rates than females across the rearrest categories of interest. While the failure rates were not consistently and significantly different, there were enough trends noted that future research should monitor these differences.

Overall, when considering six different pretrial outcome measures, the PTRA performs in the "good" or "excellent" range and provides meaningful categorizations. When considering the two outcomes that are likely the least subjective (i.e., rearrest for any offense and rearrest for a violent offense), there are no consistently significant differences in the performance of the PTRA across matched groups based on race, ethnicity, or sex. While both Hispanics and females seemed to have lower failure rates, these differences did not appear to be consistently significantly different as they depended upon the rearrest outcomes being examined. Even so, future research should monitor these trends, with appropriate changes in the text describing failure rates being added for Hispanic and/or female offenders if necessary.

The extant research, moreover, points to several potential approaches for modifying the PTRA's use in the federal pretrial system. First, it might be advisable to reformulate this instrument so that the risk categories are collapsed from five to four groups. The finding of no difference in the any rearrest rate for non-Hispanic Whites with PTRA 4 and 5 risk scores augurs for merging these risk categories. The possibility of using only four risk classification groupings, however, should be tempered by the fact that the violent rearrest rates for non-Hispanic White defendants do differ across the PTRA 4 and 5 categories; moreover, PTRA 5 Blacks and Hispanics have higher rearrest rates than their PTRA 4 counterparts. Hence, more research and testing of whether fewer risk categories could provide a better fit with the data would be advisable before initiating this potential change. Another possible modification involves generating separate pretrial violation tables by ethnicity and sex in the AOUSC's case management system for pretrial officers to review. Given the mixed evidence of overprediction for females and Hispanics, highlighting the differential pretrial violation rates for these subpopulations could address the predictive bias issue by allowing officers to observe and note each group's specific rearrest rates by PTRA risk category. At the postconviction stage, for example, the AOUSC's case management system produces tables highlighting the rearrest rates for males and females by risk classification group.

It should be clearly understood that this discussion is limited to the performance of the PTRA as an instrument. The fact that there is no consistent evidence of test bias across race, ethnicity, and sex means that the instrument generally performs the same for these groups. Stated differently, the instrument achieves predictive parity and for the most part is well calibrated (Chouldechova, 2016). These qualities, however, do not necessarily translate into "fair" decisions. Just as Skeem and Lowenkamp (2016) noted regarding the use of a calibrated risk assessment in postconviction settings, making decisions based on the PTRA might or might not lead to disparate treatment. Fairness in decision making can be altogether different than fairness in assessment performance (Corbett-Davies et al., 2017). As such, future research should focus on how the most consistent and fair decisions can be made using a pretrial risk assessment. This might require special attention to cutoffs for creating risk bins as well as the development of decision rules regarding pretrial detention and alternatives to pretrial detention (e.g., third-party custody, electronic monitoring) that might be seen, by some, as better than detention but restrictive nonetheless. Finally, future research might be directed at using emerging statistical tools and methodologies including machine learning techniques to

further enhance this tool's risk prediction capacities for defendants in general and across the race, ethnic, and sex subgroups investigated in this study.

CONCLUSION

The current study sought to examine the PTRA's capacity to predict pretrial violations among federal defendants and to investigate the instrument for predictive biases across defendant demographic characteristics. Findings from this research show that the PTRA performs well predicting various forms of pretrial violations, including rearrests for any or violent offenses, FTAs, pretrial revocations, or a combination of these outcomes. This finding supports the contention that officers can use the PTRA to gauge a defendant's likelihood of committing pretrial recidivism and various other forms of pretrial violation activity and hence apply this instrument when making release recommendations. Moreover, this research demonstrates that the PTRA can predict violations irrespective of the defendant's race, ethnicity, and sex. These findings are supportive of a growing literature showing that risk instruments like the PTRA can be used to assess recidivism risk and inform criminal justice decisions without exacerbating biases in the criminal justice system (Skeem & Lowenkamp, 2016; Skeem et al., 2016). Although we have shown how well the PTRA predicts, this research has not explored whether, and the extent to which, decisions based on the PTRA might be leading to race, ethnic, or sex-based disparities. Subsequent research might contemplate moving beyond the issue of risk prediction and focus on how decision makers actually use information generated from actuarial risk instruments to inform their decisions, policies, and practices.

NOTES

1. In this study, the term race bias refers to whether the pretrial risk assessment instrument (PTRA) predicts pretrial violation outcomes equally well between non-Hispanic Whites and Blacks, while ethnic bias denotes an examination of the PTRA's predictive efficacy for non-Hispanic Whites and Hispanics of any race. Sex bias refers to the biological attributes of men and women rather than the social construction characteristics attributed by gender; hence, this component of the study will examine the PTRA's predictive calibration between men and women.

2. Pretrial recidivism means that the defendant was arrested for new criminal conduct. It does not refer to missed court appearances or revocations from pretrial release.

REFERENCES

American Bar Association. (2007). *ABA standards for criminal justice: Pretrial release* (3rd ed.). Washington, DC: American Bar Association.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: Matthew Bender & Company.

Angwin, J., Larson, M., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Austin, J., Bhati, A., Jones, M., & Ocker, R. (2010). *Florida pretrial risk assessment instrument*. Denver, CO: The JFA Institute.

Babchishin, K. M., & Helmus, M. L. (2016). The influence of base rates on correlations: An evaluation of proposed alternative effect sizes with real-world data. *Behavior Research Methods*, *48*, 1021-1031.

Baber, L. (2010). Results-based framework for post-conviction supervision recidivism analysis. *Federal Probation*, *74*(3), 5-10.

Bechtel, K., Lowenkamp, C. T., & Holsinger, A. (2011). Identifying pretrial failure: A meta-analysis. *Federal Probation*, *75*(2), 78-87.

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2017). Fairness in criminal justice risk assessments: The state of the art. *arXiv*. Retrieved from https://arxiv.org/abs/1703.09207

Bersani, B. E., & Piquero, A. R. (2016). Examining systematic crime reporting bias across three immigrant generations: Prevalence, trends, and divergence in self-reported and official reported arrests. *Journal of Quantitative Criminology*, *33*, 835-857.

Blomberg, T., Bales, W., Mann, K., Meldrum, R., & Nedelec, J. (2010). *Validation of the COMPAS risk assessment classification instrument*. Tallahassee, FL: College of Criminology and Criminal Justice Florida State University.

Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior*, *36*, 21-40.

Cadigan, T. P., Johnson, J. L., & Lowenkamp, C. T. (2012). The re-validation of the federal pretrial services risk assessment (PTRA). *Federal Probation*, *76*(2), 3-9.

Cadigan, T. P., & Lowenkamp, C. T. (2011). Implementing risk assessment in the federal pretrial system. *Federal Probation*, *75*(2), 30-34.

Chouldechova, A. (2016). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *arXiv*. Retrieved from https://arxiv.org/abs/1610.07524

Chouldechova, A., & G'Sell, W. (2017). Fairer and more accurate, but for whom? *arXiv*. Retrieved from https://arxiv.org/abs/1707.00046

Cohen, T. H. (2013). *Pretrial detention and misconduct in federal district courts, 1995-2010*. Washington, DC: Bureau of Justice Statistics.

Cohen, T. H., & Reaves, B. A. (2007). *Pretrial release of felony defendants in state courts*. Washington, DC: Bureau of Justice Statistics.

Connolly, M. (2003). *A critical examination of actuarial offender-based prediction assessments: Guidance for the next generation of assessments*. Washington, DC: National Institute of Justice.

Cook, D. (2015). *CCMATCH: Stata module to randomly match cases and controls based on specified criteria*. Retrieved from http://www.Danielecook.com

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. *arXiv*. Retrieved from https://arxiv.org/abs/1701.08230

Criminal Justice Policy Program. (2016). *Moving beyond money: A primer on bail reform*. Cambridge, MA: Criminal Justice Policy Program Harvard Law School.

Danner, M. J. E., VanNostrand, M., & Spruance, L. M. (2016). *Race and gender neutral pretrial risk assessment, release recommendations, and supervision: VPRAI and PRAXIS revised*. St. Petersburg, FL: Luminosity.

Desmarais, S. L., & Singh, J. (2013). *Risk assessment instruments validated and implemented in correctional settings in the United States*. New York, NY: Council of State Governments Justice Center.

Dieterich, W., Mendoza, C., & Brennan, T. (2016). *COMPAS risk scales: Demonstrating accuracy equity and predictive parity*. Ottawa, Ontario, Canada: Northpointe Research Department.

Eskridge, C. W. (1983). *Pretrial release programming: Issues and trends*. New York, NY: Clark Boardman Company.

Flores, A. W., Bechtel, K., & Lowenkamp, C. T. (2016). False positives, false negatives, and false analyses: A rejoinder to "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks." *Federal Probation*, *80*(2), 38-46.

Goldkamp, J. S. (1985). Danger and detention: A second generation of bail reform. *Journal of Criminal Law and Criminology*, *76*(1), 1-74.

Gottfredson, M. R., & Gottfredson, D. (1988). *Decision making in criminal justice: Toward the rational exercise of discretion* (2nd ed.). New York, NY: Plenum Press.

Gupta, A., Hansman, C., & Frenchman, E. (2016). The heavy costs of high bail: Evidence from judge randomization. *Journal of Legal Studies*, *45*, 471-505.

Hamilton, M. (2015). Risk-needs assessment: Constitutional and ethical challenges. *American Criminal Law Review*, *52*, 231-291.

Hanson, R. K. (2017). Assessing the calibration of actuarial risk scales: A primer on the E/O index. *Criminal Justice and Behavior*, *44*, 26-39.

Harcourt, B. (2008). *Against prediction: Profiling, policing, and punishing in an actuarial age*. Chicago, IL: The University of Chicago Press.

Harcourt, B. (2015). Risk as a proxy for race: The dangers of risk assessment. *Federal Sentencing Reporter*, *27*, 237-243.

Harrington, M., & Spohn, C. (2007). Do race, age, and sex matter in sentencing decisions? Use of the total incarceration variable in a non-guideline jurisdiction. *Journal of Research in Crime and Delinquency*, *44*, 36-63.

Hart, T., & Reaves, B. A. (1999). *Felony defendants in large urban counties, 1996*. Washington, DC: Bureau of Justice Statistics.

Heaton, P., Mayson, S. G., & Stevenson, M. (2017). The downstream consequences of misdemeanor pretrial detention. *Stanford Law Review*, *69*, 711-794.

Hilbe, J. (2009). *Logistic regression models*. New York, NY: CRC Press.

Hoge, R. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior*, *29*, 380-396.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv*. Retrieved from https://arxiv.org/abs/1609.05807

Klingele, C. (2016). The promises and the perils of evidence-based corrections. *Notre Dame Law Review*, *91*, 537-576.

Latessa, E. J., Lemke, R., Makarios, M., Smith, P., & Lowenkamp, C. T. (2010). The creation and validation of the Ohio Risk Assessment System (ORAS). *Federal Probation*, *74*(1), 16-22.

Laura and John Arnold Foundation. (2013). *Developing a national model for pretrial risk assessment*. Houston, TX: Author.

Laura and John Arnold Foundation. (2016). *Public safety assessment: Risk factors and formula*. Houston, TX: Author.

Levin, D. J. (2010). *Validation of the Coconino County pretrial risk assessment tool*. Rockville, MD: Pretrial Justice Institute.

Lowenkamp, C. T., & Bechtel, K. (2007). Predictive validity of the LSI-R on a sample of offenders drawn from the records of the Iowa Department of Correction Data Management System. *Federal Probation*, *71*(3), 25-34.

Lowenkamp, C. T., Holsinger, A., & Cohen, T. H. (2015). PCRA revisited: Testing the validity of the federal Post Conviction Risk Assessment (PCRA). *Psychological Services*, *12*, 149-157.

Lowenkamp, C. T., VanNostrand, M., & Holsinger, A. (2013). *Investigating the impact of pretrial detention on sentencing outcomes*. Houston, TX: Laura and John Arnold Foundation.

Lowenkamp, C. T., & Whetzel, J. (2009). The development of an actuarial risk assessment instrument for U.S. Pretrial Services. *Federal Probation*, *73*(2), 33-36.

Mamalian, C. A. (2011). *State of the science of pretrial risk assessment*. Washington, DC: Bureau of Justice Assistance.

Mayson, S. G. (2018). Bail reform and restraint for dangerousness: Are defendants a special case? *Yale Law Journal*, *127*, 490-569.

Monahan, J., Skeem, J., & Lowenkamp, C. T. (2017). Age, risk assessment, and sanctioning: Overestimating the old, underestimating the young. *Law and Human Behavior*, *41*, 191-201.

Oleson, J. C., VanNostrand, M., Lowenkamp, C. T., Cadigan, T. P., & Wooldredge, J. (2014). Pretrial detention choices and federal sentencing. *Federal Probation*, *78*(1), 12-18.

Pepin, A. (2013). *Evidence-based pretrial release*. Williamsburg, VA: National Center for State Courts.

Piquero, A. R., & Brame, R. (2008). Assessing the race–crime and ethnicity–crime relationship in a sample of serious adolescent delinquents. *Crime & Delinquency*, *54*, 390-422.

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

Silver, E., & Miller, L. (2002). A cautionary note on the use of actuarial risk assessment tools for social control. *Crime & Delinquency*, *48*, 138-161.

Skeem, J., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. *Criminology*, *54*, 680-712.

Skeem, J., Monahan, J., & Lowenkamp, C. T. (2016). Gender, risk assessment, and sanctioning: The cost of treating women like men. *Law and Human Behavior*, *40*, 580-593.

Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology and Public Policy*, *8*, 183-208.

Starr, S. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review*, *66*, 803-872.

Starr, S. (2015). The new profiling: Why punishing based on poverty and identity is unconstitutional and wrong. *Federal Sentencing Reporter*, *27*, 229-236.

Stevenson, M., & Mayson, S. G. (2017). Bail reform: New directions for pretrial detention and release. *Faculty Scholarship*. Retrieved from https://scholarship.law.upenn.edu/faculty_scholarship/1745

Stevenson, M. (in press). Assessing risk assessment in action. *Minnesota Law Review*.

Tartaro, C., & Sedelmaier, C. M. (2009). A tale of two counties: The impact of pretrial release, race, and ethnicity upon sentencing decisions. *Criminal Justice Studies: A Critical Journal of Crime, Law and Society*, *22*, 203-221.

Ulmer, J. T. (2012). Recent developments and new directions in sentencing research. *Justice Quarterly*, *29*, 1-40.

VanNostrand, M., & Keebler, G. (2009). Pretrial risk assessment in the federal court. *Federal Probation*, *73*(2), 3-29.

Van Voorhis, P., & Brown, K. (1996). *Risk classification in the 1990s*. Washington, DC: U.S. Department of Justice, National Institute of Corrections.

Vaughn, M. G., Salas-Wright, C. P., DeLisi, M., & Maynard, B. R. (2014). The immigrant paradox: Immigrants are less antisocial than native-born Americans. *Social Psychiatry and Psychiatric Epidemiology*, *49*, 1129-1137.

Thomas H. Cohen received his PhD in criminal justice from the University of Rutgers School of Criminal Justice in 2003 and his JD from the University of Maryland School of Law in 1996. Currently, Dr. Cohen is a social science analyst at the Administrative Office of the United States Courts (AOUSC), Probation and Pretrial Services Office. His work includes analyzing risk assessment data at the postconviction and pretrial levels and authoring reports on how the AOUSC integrates the risk principle into its operational practices. His recent research has appeared in *Criminology and Public Policy*, *Criminal Justice Policy Review*, *Federal Probation*, *Journal of Empirical Legal Studies*, and *Psychological Services*. Moreover, he has

authored several technical reports on criminal court case processing at the state and federal levels through his prior work at the Bureau of Justice Statistics.

Christopher T. Lowenkamp received his PhD in criminal justice from the University of Cincinnati. He has served as the director of the Center for Criminal Justice Research, the associate director of The Corrections Institute at the University of Cincinnati, and research professor at the University of Cincinnati. He is currently a social science analyst for the Administrative Office of the United States Courts (AOUSC). He has co-authored more than 65 articles and book chapters, some of which are published in top-tier academic journals.