DOI: 10.1111/jels.12343

ORIGINAL ARTICLE

JOURNAL OF EMPIRICAL LEGAL STUDIES

Racial bias as a multi-stage, multi-actor problem: An analysis of pretrial detention

Joshua Grossman[1] | **Julian Nyarko**[2] | **Sharad Goel**[3]

[1]Stanford University, Stanford, California, USA

[2]Stanford Law School, Stanford, California, USA

[3]Harvard University, Cambridge, Massachusetts, USA

Correspondence
Joshua Grossman, Department of Management Science and Engineering, Stanford University, 475 Via Ortega, Stanford, CA 94305, USA.
Email: jdgg@stanford.edu

Abstract

After arrest, criminal defendants are often detained before trial to mitigate potential risks to public safety. There is widespread concern, however, that detention decisions are biased against racial minorities. When assessing potential racial discrimination in pretrial detention, past studies have typically worked to quantify the extent to which the ultimate judicial decision is conditioned on the defendant's race. Although often useful, this approach suffers from three important limitations. First, it ignores the multi-stage nature of the pretrial process, in which decisions and recommendations are made over multiple court appearances that influence the final judgment. Second, it does not consider the multiple actors involved, including prosecutors, defense attorneys, and judges, each of whom have different responsibilities and incentives. Finally, a narrow focus on disparate *treatment* fails to consider potential disparate *impact* arising from facially neutral policies and practices. Addressing these limitations, here we present a framework for quantifying disparate impact in multi-stage, multi-actor settings, illustrating our approach using 10 years of data on pretrial decisions from a federal district court. We find that Hispanic defendants are released at lower rates than white defendants of similar safety and nonappearance risk. We trace these disparities to decisions of assistant US attorneys at the initial hearings, decisions driven in part by a statutory mandate that lowers the procedural bar for moving for detention of defendants in certain types of cases. We also find that the Pretrial Services Agency recommends detention of Black defendants at higher rates than white defendants of similar risk, though we do not find evidence that these recommendations translate to disparities in actual release rates. Finally, we find that traditional disparate treatment analyses yield more modest evidence of discrimination in pretrial detention outcomes, highlighting the value of our more expansive analysis for identifying, and ultimately remediating, unjust disparities in the pretrial process. We conclude with a discussion of how risk-based threshold release policies could help to mitigate observed disparities, and the estimated impact of various policies on violation rates in the partner jurisdiction.

INTRODUCTION

An extensive body of work investigates racial disparities in the criminal justice system and its causes, and there is ample evidence of discrimination in different parts of the process, including policing and arrests (Antonovics & Knight, 2009; Fryer, 2019; Pierson et al., 2020), charging (Chohlas-Wood et al., 2021; Rehavi & Starr, 2014), bail and pretrial detention (Arnold et al., 2018, 2022; Grau & Vergara, 2020; Kutateladze et al., 2012), plea bargaining (Metcalfe & Chiricos, 2018), conviction and sentencing (Anwar et al., 2012), and incarceration (Abrams et al., 2012). Among the different decision points in the criminal process, whether or not to detain a defendant until their trial is of particular importance, in part because it can significantly influence downstream outcomes. Evidence suggests that defendants subjected to pretrial detention are more likely to plead guilty (Leslie & Pope, 2017; Sutton, 2013) and receive harsher sentences (Didwania, 2020; Dobbie et al., 2018; Heaton et al., 2017; Oleson et al., 2016; Spohn, 2008; Stevenson, 2018; Sutton, 2013). Furthermore, when found guilty, defendants detained pretrial may be more likely to recidivate following their sentence (Heaton et al., 2017; Lowenkamp et al., 2013). Meanwhile, defendants released pretrial can engage in programs and activities giving them the opportunity to demonstrate that a shorter or noncustodial sentence is warranted, in turn reducing costs imposed by longer sentences on defendants, their communities, and the carceral system (Carr, 2016; Oleson et al., 2016).

Although previous studies have examined discrimination in the pretrial process (Arnold et al., 2018, 2022; Ayres & Waldfogel, 1993; Demuth, 2003; Demuth & Steffensmeier, 2004; Grau & Vergara, 2020; Hull, 2021; Lynch & Patterson, 1991; Schlesinger, 2005, 2007; Spohn, 2008; Sutton, 2013; Turner & Johnson, 2007), they are subject to significant shortcomings. First, past studies have conceptualized pretrial detention as the result of a unitary decision process involving only a judge. But pretrial detention, like other outcomes in the criminal justice system, results from an interrelated series of decisions, and studying one decision point does not always capture the root causes of discrimination, as disparities may accrue or subside as a case proceeds from one stage to the next (Schlesinger, 2007). For instance, the typical federal pretrial process consists of at least three decision points with direct carceral consequences for the defendant. First, there is an initial hearing during which it is determined whether the defendant is "held for detention" or released. Then, if held for detention, there is a detention hearing to determine whether the defendant will be detained. Finally, between hearings, defendants have the possibility to consent to detention.

Given these complexities, one might be inclined to follow the prescriptions of a more recent strand of the literature and focus on cumulative disparities (Arnold et al., 2018, 2022; Kurlychek & Johnson, 2019; Kutateladze et al., 2014; Omori & Johnson, 2019; Stolzenberg et al., 2013; Sutton, 2013). But an exclusive emphasis on cumulative disparities, too, can mask important costs for the criminal defendant. For instance, even though a cumulative analysis may, hypothetically, suggest that there are no disparities in final detention decisions, there could still be disparities at the initial hearing, with minority defendants spending additional time in jail. Given that costs are disproportionately borne at the very beginning of incarceration, such a pattern would mean that marginalized groups face significant additional burdens that are not reflected in analyses that are limited to the final outcome of the pretrial process. Furthermore, by examining aggregate disparities across the process as a whole, a cumulative analysis makes it difficult to design targeted interventions to remediate problems at specific stages. Thus, instead of embracing either the individualized or the cumulative view, it appears appropriate to examine disparities both at individual decision points and cumulatively across decision points.

A second way in which the pretrial detention process defies typical assumptions made in the literature concerns the number of actors involved in the decision-making process. Rather than being an isolated decision of a judge, pretrial detention decisions result from a complex interplay between at least three actors in any case: (1) an assistant US attorney (AUSA), who decides whether to move for detention; (2) a pretrial services (PTS) officer, who prepares a bail report and recommends detention or release; and (3) a judge, who makes the detention decision subject to certain statutory constraints. This interplay suggests that disparities can enter the detention process not only at separate decision points, but also through several different actions taken by numerous actors (Bohren et al., 2022). In order to formulate effective policy proposals aimed at reducing disparities, it is important to take these nuances into consideration and to identify the concrete source of disparities. At the same time, obtaining data that allow for a detailed analysis of relevant decision makers can be difficult, given that information on AUSA motions for detention, PTS bail reports, and intermediate judicial decisions are often not publicly available.

Finally, a third shortcoming of past studies of pretrial detention is that, with some notable exceptions (Arnold et al., 2018, 2021, 2022), they have taken a narrow view of what constitutes discriminatory conduct.[1] In particular, they employ a methodology that seeks to assess whether detention decisions are implicitly or explicitly conditioned on race. The primary statistical concern of

[1] The Arnold et al. (2021) measure of disparate impact is based on outcomes that are only observed after a judicial decision is made. Importantly, and in contrast, our definition of risk is based on information available at the time of the decision. In our setting, and under the conceptualization of disparate impact articulated by Jung et al. (2019), it is critical to consider such ex ante measures of risk in order to determine whether similarly risky people are treated similarly. Both measures may be potentially useful ways to characterize disparate impact, but we believe that the approach we take is more closely aligned with traditional legal doctrine. In particular, courts have deemed legally permissible policies that first rank individuals by ex ante risk and then distribute resources (e.g., loans) to those above a threshold risk level. Such threshold policies would likewise not be found to have disparate impact under our definition but would in general violate definitions based on ex post outcomes (Corbett-Davies & Goel, 2018). Of course, aside from these legal considerations, there are larger normative concerns about how one should assess the equity of different policies, which we leave to future work.

researchers conducting such studies arises out of omitted-variable bias. Consequently, a commonly employed design involves controlling for as many observable covariates as possible, and assessing whether the residual variation in outcomes is explained by the defendants' race. This view of discrimination approximately translates into an understanding of discrimination as disparate *treatment*. Although important, that view ignores potential discriminatory disparate *impact*, which can arise from facially neutral actions that impose burdens on racial minorities without furthering a compelling policy goal.

In this paper, we aim to address these three limitations of past work, explicitly modeling pretrial detention as a multi-stage, multi-actor process, and taking an expansive view of discrimination that includes disparate impact stemming from facially neutral policies and practices. To achieve this goal, we partnered with a US district court (henceforth "our partner jurisdiction") to conduct a detailed analysis of the pretrial process in their cases involving defendants with US citizenship. Through our collaboration, we gained access to structured data on hearing dates and outcomes along with defendant demographics, charges, and criminal history. In addition to this structured data, we received access to unstructured bail reports containing additional information, such as descriptions of defendants' ability to pay bail. To make use of these reports—which differ in structure, style, and content across jurisdictions, and may even differ in format from year to year in the same jurisdiction—we developed an automated process to anonymize and extract the relevant information. In this way, our study is the first to analyze disparities at a previously unattainable level of granularity, allowing us to determine specific decision points and actors that give rise to disparities.

Importantly, our study expands upon the scope of past analyses and considers not only whether detention decisions are—either implicitly or explicitly—conditioned on race, but rather whether observed disparities are fully explained by justified, risk-related factors known to the individual decision maker. We thus adopt a view of discriminatory conduct that is consistent with the doctrine of disparate *impact*. Although disparate impact by government actors is outlawed only under narrow circumstances, our empirical approach builds on the rationale that the entrenchment of racially disparate outcomes should be avoided if the creation of such disparities does not further a legitimate policy goal. In the context of pretrial decision making, the main justifications to detain a defendant who has not yet been found guilty of a crime are a concern for public safety and the risk of nonappearance at a future court date (c.f. 18 U.S. Code § 3142 (c)). We thus empirically assess whether disparities in detention decisions between Black, Hispanic, and white[2] defendants stem (fully) by observable risk factors or whether there are residual disparities that cannot be explained by the stated policy objectives.

[2] Courts typically refer to non-Hispanic white defendants as "white" and Hispanic white defendants as "Hispanic." For consistency, we adopt those labels.

We find that Hispanic defendants are released at the initial hearing at lower rates than white defendants of similar safety and nonappearance risk, with directional—though not statistically significant—evidence of a similar disparity for Black defendants. This result is primarily driven by risk-adjusted disparities in the rates at which the AUSA moves for detention of Black and Hispanic defendants at the initial hearing, a pattern that appears to be driven in part by a statutory mandate entitling the AUSA to a detention hearing in certain types of cases. Between the initial hearing and the detention hearing, PTS investigates the defendant's fitness for release and prepares a detention or release recommendation for the judge. Even though PTS recommends detention for Black defendants at higher rates than comparably risky white defendants, we do not observe risk-adjusted disparities in the rates at which Black and white defendants are released after the PTS investigation. Overall, we find that Hispanic defendants are less likely to be released at any point in the pretrial process than similarly risky white defendants, with statistically insignificant estimates of risk-adjusted disparities for Black defendants.

As a point of comparison, we find that a traditional disparate treatment analysis of the ultimate detention outcome—in which one adjusts for all observable factors, regardless of their relevance to policy goals—yields far more modest differences in release rates for Hispanic versus white defendants. Our analysis thus shows that treating the pretrial process as a unitary decision-making process, and focusing narrowly on disparate treatment, risks masking significant disparities that produce unjustified burdens on marginalized groups. In addition, our results show that not all disparities induced by individual actors have the same consequence. In particular, the evidence suggests that judges can act as important gatekeepers able to successfully counteract disparities in PTS recommendations at the detention hearing. However, in order to serve as an effective check, judges need to acquire a minimum level of information sufficient to overturn the recommendation provided to them. At the initial hearing, they typically have little information about the idiosyncratic circumstances of individual defendants, and so may be more willing to accept motions for detention made by the AUSA. This information asymmetry could help explain why disparities in motions for detention of defendants at the initial hearing translate into disparities in the actual decisions at the initial hearing, whereas disparities in recommendations provided at the detention hearing are largely not reflected in the corresponding decisions at the detention hearing.

We highlight two potential pathways to mitigating racial disparities in the partner jurisdiction. First, given the lack of observed disparities in the later stages of the pretrial process, it appears plausible that earlier access to estimates of the defendant's pretrial risk could help to mitigate disparities in the decisions of AUSAs at the initial hearing. Many other federal jurisdictions already conduct PTS investigations before the initial hearing, providing AUSAs and judges with access to estimates of safety and nonappearance risk at the initial hearing. Second, we propose the implementation of a risk-based threshold release policy. Under that policy, all defendants below a particular risk score threshold are presumptively released. Historically, defendants in our partner jurisdiction violate the terms of release at modest rates. In particular, out of all released defendants in our sample, 2% failed to appear at a mandatory court date, 2% were arrested for any offense, less than 1% were arrested for a felony offense, and 6% committed a violation serious enough to warrant a revocation of release. Low violation rates raise the possibility that more defendants could be released with only a modest increase in violations.

A BRIEF PRIMER ON THE FEDERAL PRETRIAL PROCESS

The federal pretrial process involves a complex set of procedures that span multiple decision points and involve several different actors making both decisions and recommendations that influence the final outcome. Figure 1 outlines the key steps in this process, which we also briefly discuss below. We describe the process followed by our partner jurisdiction, though there are broad similarities across districts.

After a defendant is summoned to court, or within 48 hours of their arrest, the court holds an *initial hearing*. At this point, the AUSA decides whether or not to move for a *detention hearing*. 18 U.S. Code § 3142 (f) defines certain charges for which the AUSA is *entitled* to a detention hearing. These include violent offenses, drug or terrorism offenses with a maximum term of 10 years or more, offenses that carry a maximum sentence of life imprisonment or death, any felony offense if the defendant has been convicted of two or more of the aforementioned offenses, and any nonviolent felony that involves a dangerous weapon, a minor victim, or a failure to register as a sex offender. If at least one of the defendant's charges falls under the entitlement provision and the AUSA moves for detention, the judge has no discretion and the defendant will be "held for detention" (i.e., they will be held until their detention hearing takes place). If none of the charges are subject to the entitlement provision, the AUSA can still argue for a detention hearing if it is shown that the defendant presents a serious risk of flight or obstruction of justice. As a practical matter, if the AUSA does not move for detention, the defendant is typically released. If a defendant is released, it is either on a bond, or on their own recognizance (i.e., without supervision by the Federal Pretrial Services Agency [PTS]).

If scheduled, the detention hearing typically takes place within three business days of the initial hearing. Between the hearings, PTS conducts an investigation to determine the defendant's fitness for release, and, if deemed fit, the release

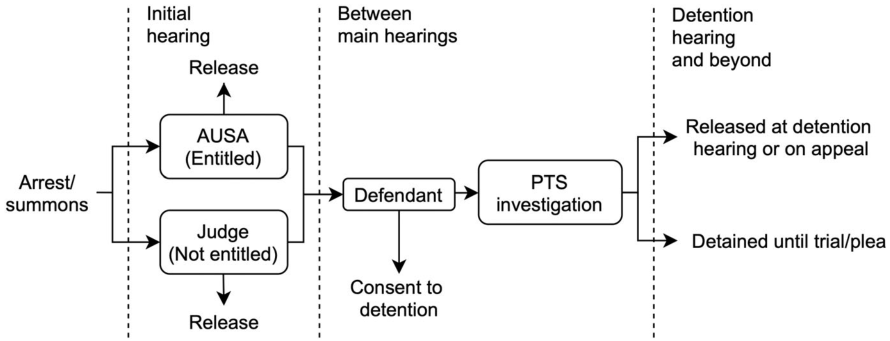

FIGURE 1 Key pretrial decision points. Shortly after a defendant is arrested or summoned to court, the court holds an initial hearing. If at least one of the defendant's charges triggers the entitlement to a detention hearing, the assistant US attorney (AUSA) decides whether to proceed with a detention hearing ("move for detention") or allow the defendant to be released. If no charges trigger the entitlement to a detention hearing and the AUSA moves for detention, the judge decides whether there is sufficient reason to hold a detention hearing, and either releases the defendant or schedules a detention hearing. At any point in the pretrial process, defendants may choose to consent to detention. If a defendant does not consent to detention just after the initial hearing, PTS conducts its investigation of the defendant and formulates a recommendation for detention or release. If a detention hearing is held, the judge weighs arguments from the AUSA and the defense, along with the recommendation from PTS, to decide whether to release or "preventively detain" the defendant before trial. Finally, if detained at the detention hearing, defendants are permitted to appeal for release before their trial or plea agreement. Figure A1 is an expanded version of this figure.

conditions required to minimize nonappearance and public safety risk. Defendants may choose to waive their right to a detention hearing ("consent to detention") before, during, or after the PTS investigation, and defendants consent to detention for a variety of reasons.[3] For example, they may need additional time to prepare a stronger case for release, or they may want to avoid the scrutiny inherent to the PTS investigation, including having their friends and family interviewed as potential bail resources. Other defendants may not want to openly discuss the contents of the PTS recommendation during a detention hearing. At any point after consenting to detention, new information may become available that affects the fitness of the defendant for pretrial release. For example, while a friend or family member may be initially unwilling to serve as a bail resource, they are free to change their mind at any point after the

[3] Undocumented defendants who suspect that they will eventually be found guilty may choose to waive their right to a detention hearing even if they could otherwise be released pretrial, as time spent in Immigration and Customs Enforcement (ICE) custody does not count towards time served, and detention facilities contracted by the US Marshals Service typically have better amenities and services than ICE facilities. We exclude undocumented defendants from the analysis.

defendant consents to detention. If the friend or family member later decides to serve as a bail resource, bail proceedings may be reopened.

If a defendant is not released at the initial hearing and does not consent to detention, the court holds a detention hearing. At the detention hearing, the judge weighs arguments from the AUSA and the defense, along with the PTS recommendation, in deciding whether to release or "preventively detain" the defendant. If the defendant is detained at the detention hearing, they may appeal for release at any point before their trial or plea.

The above description of the federal pretrial process is accurate for the vast majority of cases in our partner jurisdiction, though occasionally defendants are released or detained through other means. For example, while very rare, defendants may be preventively detained at the initial hearing.

DISENTANGLING DISPARATE TREATMENT AND DISPARATE IMPACT

US law recognizes two distinct discrimination doctrines: disparate *treatment* and disparate *impact*. Under disparate treatment, an action or policy is considered discriminatory if it is motivated by a discriminatory purpose or animus toward the protected group. The Equal Protection Clause of the Fourteenth Amendment outlaws disparate treatment of racial minorities by government actors, unless the government can substantiate a compelling interest.[4] In contrast, under the doctrine of disparate *impact*, a policy or action is discriminatory not because it is conditioned on race, but because it produces unjustified, disparate outcomes to the disadvantage of the protected group. Policies that produce a disparate impact are not generally outlawed under the US Constitution. Instead, federal laws like the Civil Rights Act of 1964 and different state laws render both public and private policies producing a disparate impact illegal in certain domains, such as employment, credit, and housing. In the criminal justice context, there is no broad federal prohibition against disparate impact, though some states, including California and Illinois, have extended disparate impact protections to this domain.

Although they differ in scope and coverage, both disparate treatment and disparate impact doctrine seek to protect against conduct that is normatively fraught. A holistic empirical assessment of potentially problematic practices—irrespective of their legality—thus requires considering both types of discrimination, each with its own methodological approach. Most studies examining disparities in the criminal justice context estimate a regression model of the following or similar functional form (Gaebler et al., 2022):

[4] Such as under certain affirmative action programs (Fisher v. U of Texas 2016) or to remedy past discrimination (Wygant v. Jackson Bd. of Educ. 1986).

$$\Pr(Y_i = 1) = \text{logit}^{-1}\left(\alpha_{\text{race}[i]} + \beta^T \vec{X}_i\right), \tag{1}$$

where Y_i is the binary outcome of the decision, $\alpha_{\text{race}[i]}$ is an intercept term shared by defendants with the same race or ethnicity as defendant i, and $\vec{X}_i$ is a vector of additional controls. This estimation strategy is motivated by a desire to estimate the causal effect of race—or perceptions thereof (Greiner & Rubin, 2011)—on decisions.[5] Implicitly, this design embraces a narrow definition of discrimination as disparate *treatment*: the researcher wants to know, for example, whether a Black defendant is treated differently from a white defendant *because of* their race. For instance, the researcher may be interested in $\alpha_{\text{Black}} - \alpha_{\text{white}}$ as a measure of the racial gap in decisions among similarly situated individuals. The primary statistical concern in these studies is omitted-variable bias. Hence, it is typical for studies in this setting to include as many observable controls as possible in $\vec{X}_i$. By adjusting for a large number of factors, the hope is that the design allows for the conclusion that differences in outcomes can be traced back to differences in the defendants' race as opposed to other dimensions, such as criminal record or socioeconomic status.

A disparate *impact* analysis necessitates a different type of regression. To see why, recall that disparate treatment involves racial disparities in outcomes among individuals who are otherwise identical, whereas disparate impact involves disparities in outcomes that are unjustified by policy goals. For instance, the main justification for detaining a defendant until their trial is to mitigate risks to public safety and to ensure appearance at required court dates. Assuming that these risks can be observed or estimated, under a disparate impact analysis one should examine racial differences in outcomes among similarly risky individuals. In particular, the analyst should not control for socio-economic factors in addition to a defendant's (estimated) risk. Although some socio-economic factors might be predictive of the expected harm a defendant causes to society after release, the defendant's risk variables would already capture all policy-relevant information contained in the socioeconomic factors. By explicitly including these socioeconomic factors in addition to risk covariates, the researcher would control for defendant characteristics above and beyond the extent justified by the policy goal that the detention process tries to achieve. One central insight flowing from this description is that, while in a disparate treatment analysis it is desirable to control for as many observables as possible,[6] in a disparate impact analysis, including too many covariates threatens controlling for unjustified drivers of disparities—in turn erroneously reducing the

[5] To be sure, many studies eschew an explicit causal framework, perhaps in hopes that the avoidance of causal language lowers the burden of proof required to demonstrate disparate treatment. However, the design choice makes it clear that the estimation of the causal effect is the primary goal.

[6] There are some caveats. For instance, one should be mindful not to control for covariates considered to be post-treatment, or those that can induce other forms of biases, such as M-biases (Ding & Miratrix, 2015; Pearl, 2009).

estimated magnitude of observed disparities, a phenomenon known as "included-variable bias" (Ayres, 2005; Jung et al., 2019).

More formally, following Jung et al. (2019), in a disparate impact analysis of pretrial detention, the model that is estimated is of the form:

$$\Pr(Y_i = 1) = \text{logit}^{-1}\left(\alpha_{\text{race}[i]} + \gamma \cdot \text{risk}_i\right), \tag{2}$$

where risk_i is the estimated public safety and/or nonappearance risk of defendant i, and γ is its associated coefficient. In this specification, $\alpha_{\text{Black}} - \alpha_{\text{white}}$, for instance, quantifies the disparate impact of the given decision on Black defendants compared to similarly risky white defendants.

Of course, one major difficulty in fitting the above equation is the task of accurately estimating a defendant's pretrial risk (risk_i). In the pretrial detention setting, an "ideal" measure of risk would result from estimating true violation behavior based on all attributes available to the decision maker at the time of the decision. In practice, violation outcomes are not perfectly observed. For example, arrests are not the same as actual criminal acts, so a risk measure that is most accurate on recorded outcomes may not be the most accurate on true outcomes. Furthermore, risk estimation can suffer from omitted variable bias through multiple channels: Researchers typically do not have access to all the information available to the decision maker at the time of the release decision, and violations can only be observed among those who are released (Lakkaraju et al., 2017). In an attempt to address these limitations, estimates of disparate impact should be examined for their robustness across multiple, albeit imperfect, measures of risk.

Our primary measure of risk is the criminal history subscore of the Federal Pretrial Risk Assessment (PTRA), an actuarial tool that estimates the public safety and nonappearance risk of federal pretrial defendants (Cadigan et al., 2012; Lowenkamp, 2009).[7] PTRA scores range from 0 to 14, with higher scores indicative of a higher risk of bond revocation, failure to appear, or rearrest. Like other actuarial tools, the PTRA was constructed by fitting a statistical model to a large sample of defendants, predicting adverse pretrial outcomes

[7] The PTRA score is the sum of two sub-scores: criminal history and "other." The criminal history score is based on prior felony convictions, prior FTAs, pending charges, offense type and class, and the defendant's age. The "other" score is based on employment status, education level, homeownership, substance abuse, and citizenship. The PTRA may still be scored if up to four of the "other" items are missing. However, in the partner jurisdiction, while the items in the criminal history section are collected for every defendant, the data for the "other" section are collected only for defendants under PTS supervision who agree to a bail interview, and is therefore not missing at random. That said, the PTRA criminal history score alone performs similarly to the full PTRA score in predicting bond revocations and failures to appear in our partner jurisdiction (Figures A4 and A5), and does not suffer from the bias induced by the missing "other" data. Thus, in our main analysis, we use the PTRA criminal history subscore to estimate non-appearance and safety risk.

based on factors such as criminal history, age, and the severity of the charged offense. In the Appendix, we confirm that our results are robust to multiple, alternative measures of risk.[8]

Risk assessment tools provide imperfect proxies of risk, though they have been found to consistently outperform criminal justice professionals and others at predicting pretrial outcomes and typically do so with less racial bias (Goel et al., 2021; Lin et al., 2020).[9] Nonetheless, it is a theoretical possibility that, among defendants with identical risk scores, officers of the court are able to accurately identify those who are more likely to have adverse pretrial outcomes—potentially providing a justification of estimated risk-adjusted disparities and constituting a key limitation of our approach.

DATA DESCRIPTION

Our partner jurisdiction provided us with detailed information on 8761 cases spanning 10 years, from October 1, 2009 to October 31, 2019.[10] The data contain several pieces of information for each case, including: (1) a list of hearings held, and, for each hearing, the date, type (initial hearing, detention hearing, violation hearing, or other hearing), and outcome (release, held for detention, preventive detention, or consent to detention); (2) the criminal charge(s)[11]; (3) the defendant's criminal history; and (4) demographic and behavioral data for each defendant, including substance use, education level, employment information, and residential status.

[8] In Tables A12–A18, we repeat the main analyses with full PTRA scores, with scores from an alternative, widely used risk assessment tool, the PSA (Advancing Pretrial Policy & Research, n.d.; Laura and John Arnold Foundation, 2013), and with adverse event probabilities estimated via an L2 (ridge) regression model trained on all released defendants tuned via *k*-fold cross validation (see Tables A2 and A3 for coefficients of this model). An adverse event is defined as a bond revocation, failure to appear, and/or rearrest. The results are qualitatively similar across risk scores.
Although the PSA is not broadly used in federal courts, the majority of the PSA training data were sourced from federal jurisdictions (DeMichele et al., 2020). Three separate risk instruments comprise the PSA: the PSA FTA (failure to appear), the PSA NCA (new criminal activity), and the PSA NVCA (new violent criminal activity). The PSA is scored entirely on age and criminal history data—none of the "other" factors in the PTRA are included. Adverse outcomes in our partner jurisdiction are predicted with similar accuracy across our chosen measures of risk (Figures A4 and A5). In Figure A6, we show that all of our measures of pretrial risk are indeed predictive of violations in our partner jurisdiction.

[9] We do not find evidence of a significantly different relationship between our chosen risk scores and violation outcomes among race and ethnicity groups (Figure A7).

[10] We denote a "case" as the proceedings for a single defendant. If there are co-defendants for the same alleged offense, each defendant is still part of their own "case," though, for nearly all cases, there are no co-defendants.

[11] Charges subject to the entitlement provision were flagged by matching charging codes to charging code descriptions provided by an expert from our partner jurisdiction. We similarly labeled charges that triggered the presumption for detention, which shifts the burden of proof of fitness for release from the prosecution to the defense (Austin, 2017).

We restricted the 8761 initial set of cases to those most relevant to our analysis. Specifically, we first restricted to the 6342 cases involving defendants with US citizenship. Thus, to our knowledge, no defendants in our sample were at risk of deportation at any point in the pretrial process. We further restricted our data to cases involving defendants recorded as Black, Hispanic, or white, as there were a limited number of individuals from other racial or ethnic groups, hindering statistical analysis. For similar reasons, we restricted our sample to defendants recorded as either male or female. These restrictions reduced our sample size to 5542. Finally, we restricted the sample to the 5208 cases in which defendants were both charged and supervised in our partner jurisdiction. These 5208 cases concern 4998 unique defendants.

A key requirement of our analysis is the accurate imputation of a defendant's public safety and nonappearance risk, as measured by the PTRA. Information to compute these scores—particularly detailed criminal history—is not always available in a structured database, and so we extracted the necessary information from the unstructured pretrial report produced by PTS for each defendant.[12] We were ultimately able to extract the necessary data to compute PTRA criminal history scores for 4920 of the 5208 above cases.

Finally, we restricted to the 4809 cases that were decided by a judge with at least 20 detention decisions in the entire dataset—as some of our statistical analysis uses judge-level fixed effects. Our primary analysis of discrimination in the pretrial process is based on these cases, and, in Table 1, we provide summary statistics for this set of cases. In Table A1, we provide

[12] Our partner jurisdiction stores criminal history data in both a structured format and in unstructured pretrial reports. However, while criminal history for every defendant is recorded in at least one unstructured pretrial report, PTS officers may choose whether to store criminal history in a structured format. When structured criminal history information is unavailable, counts of prior felonies, misdemeanors, and failures to appear are automatically recorded as zeroes, making them impossible to disentangle from true zero counts.
There are three types of pretrial reports that contain criminal history data. First, for defendants who do not consent to being interviewed by PTS, who are released at the initial hearing without PTS supervision (i.e., on their own recognizance), or who consent to detention before the PTS investigation, PTS prepares a record check report, detailing every prior charge, conviction, and failure to appear that PTS could find via a criminal records check. Second, for defendants who consent to being interviewed by PTS before the detention hearing, PTS prepares a prebail report, which typically contains the same criminal history fields from the record check reports along with the defendant's history in the community, family ties, bail resources, passport and travel status, marital status, employment status, financial resources, education, health, an assessment of the defendant's flight and safety risk, a recommendation for release or detention, and, if recommended for release, the recommended conditions of release. Third, postbail reports are prepared for defendants released at the initial hearing under the supervision of PTS who consent to being interviewed. Postbail reports contain the same information as prebail reports, with release recommendations that typically concur with the judge's release decision.
Using named-entity recognition and regular expression matching, we removed personally identifiable information in each parseable report, then extracted relevant criminal history. Next, for each case, we identified the report that best replicated the information available to the court at the time of the release or detention decision. We prioritized reports that were written closest to the date of the release or detention decision, and, if available, preferred reports written before the date of the release decision.

separate summary statistics for the released and detained defendants in our sample.[13]

RESULTS

To assess disparate treatment and disparate impact in the pretrial process, we fit regression models similar in form to those in Equations (1) and (2). However, to aid interpretation, we use linear—rather than logistic—probability models. We model each court outcome using three different sets of covariates: (1) an intercept-only model, to quantify raw disparities; (2) a disparate-impact model, adjusting for only PTRA criminal history risk scores[14]; (3) a disparate-treatment model, which controls for a variety of detailed case information, including the specific charge(s).

We use these models to examine disparities at specific decision points (e.g., the initial hearing) and in recommendations made by specific court officers (the AUSA and PTS), and similarly assess cumulative disparities across multiple stages of the process. Below we summarize our main results, and, in the Appendix, we provide a more complete description of our findings across the full set of models, decisions points, and actors. We conclude this section with an analysis of hypothetical release policies based on risk score thresholds.

Disparate treatment and disparate impact in release decisions

We start by conducting a traditional disparate treatment analysis of pretrial release outcomes. That is, we estimate the marginal effect of (perceptions of) race on being released, adjusting for all observable features, including the specific alleged offenses. We find that, among similarly situated defendants,

[13] We note that violation rates among released defendants in the jurisdiction that we study are substantially lower than those reported in some previous studies. The key distinction between our work and these previous studies is that we study a federal district court, whereas most past research has examined state and local courts. Indeed, the 2% rearrest rate in our jurisdiction is identical to the average across the entire federal docket from 2011 to 2018, and the 2% FTA rate in our jurisdiction is on par with the 1% federal docket average (Browne & Strong, 2022). It is not immediately clear why violation rates in the federal court system are so much lower than in the local courts, but we offer some possibilities suggested by officers of the court in our partner jurisdiction. First, federal courts are typically less likely than state courts to release defendants on money bail (see Browne & Strong, 2022 vs. Dobbie et al., 2018; Kleinberg et al., 2018), which can result in the detention of riskier defendants who might have been released on bail in the local courts. Second, defendants released by federal courts often have more stringent supervision requirements, such as drug treatment and electronic monitoring, both of which may lower violation rates. Finally, the population of federal defendants is older on average than those in many local court systems, again contributing to lower violation rates. In particular, released defendants are, on average, 39 years old in our partner jurisdiction, several years older than the average reported in past studies of local courts (Dobbie et al., 2018; Kleinberg et al., 2018).

[14] In Table A11, we repeat the analyses using a quadratic and cubic term for risk, finding qualitatively similar results.

TABLE 1 Summary statistics for demographic, criminal history, and pretrial process variables for the 4809 cases in our main analysis.

	All	Black	Hispanic	White
Race/ethnicity	100%	35%	25%	40%
Female	16%	17%	13%	16%
Age	37	35	33	42
PTRA score	6.5	7.9	7.0	4.9
PTRA criminal history score	3.8	4.8	4.1	2.8
Prior felonies	2.0	2.9	1.9	1.2
Prior misdemeanors	2.2	2.4	2.6	1.7
Prior FTAs	1.2	2.0	0.9	0.7
Has prior violation	46%	68%	45%	28%
Has felony charge	91%	95%	93%	85%
Has firearms charge	19%	33%	17%	8%
Has drug charge	33%	29%	48%	26%
Released at any time	59%	48%	52%	73%
Released at initial hearing	36%	25%	23%	53%
Detained at initial hearing	1%	1%	1%	1%
Consent before investigation	8%	8%	10%	7%
Released after investigation	23%	23%	28%	20%
Detained after investigation	32%	43%	37%	20%
AUSA entitled	61%	70%	74%	43%
AUSA moves for detention	63%	75%	73%	45%

Note: Means are shown for continuous variables, and proportions for binary variables. White defendants have, on average, lower risk scores and fewer priors than Black and Hispanic defendants. White defendants are also released at much higher rates, especially at the initial hearing. Finally, the AUSA is entitled to a detention hearing at higher rates for Black and Hispanic defendants, and also moves for detention of Black and Hispanic defendants at higher rates than white defendants.
Abbreviations: AUSA, assistant US attorney; FTA, failure to appear; PTRA, Federal Pretrial Risk Assessment.

Hispanic individuals are released at rates that are 4pp lower than that of white defendants (SE = 2pp), with no statistically significant differences between Black and white defendants, as indicated by the right-most hollow points in each panel of Figure 2.

However, as discussed above, this type of analysis—while common—can obscure unjustified disparate impact. To assess that possibility, we now fit a disparate impact model, adjusting only for the criminal history subscore of the PTRA. In this case, as shown in Figure 2, we find that among similarly *risky* defendants, Hispanic individuals are 6pp (SE = 2pp) less likely to be released than white individuals; release rates are comparable for white and Black defendants.

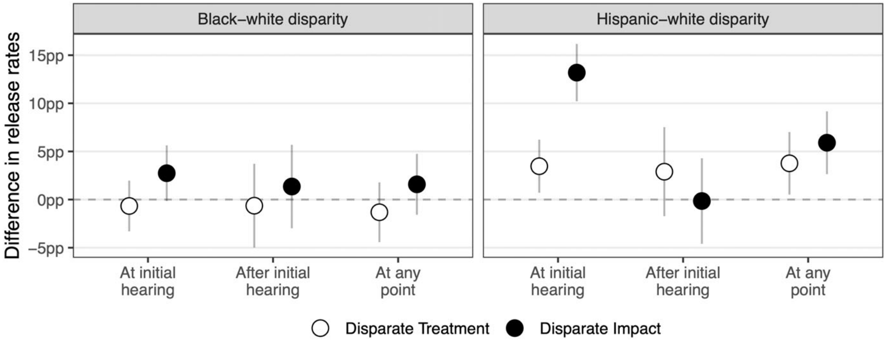

FIGURE 2 Estimated disparate impact and disparate treatment coefficients by race and release timing, with 95% confidence intervals derived from heteroskedasticity-robust standard errors. Positive values indicate discrimination towards the minority group. white defendants are more likely to be released at the initial hearing than similarly risky Black and Hispanic defendants, with a larger disparity for Hispanic defendants. The risk-adjusted disparities at the initial hearing are the primary driver of corresponding disparities in overall release rates, as we do not observe significant disparate impact estimates for release after the initial hearing. We observe significant estimates of disparate treatment of Hispanic defendants for both release at the initial hearing and release at any point. For Black defendants, disparate treatment estimates are not statistically distinguishable from zero.

To investigate the source of these disparities, we next examine release decisions at the initial hearing, and, separately, subsequent release decisions among those initially detained. As shown in Figure 2, we find a large and statistically significant 13pp risk-adjusted disparity for Hispanic defendants at the initial hearings (SE = 2pp), with a directional 3pp disparity for Black defendants (SE = 2pp). Among those initially detained, we do not see statistically significant differences in risk-adjusted release rates across groups. It thus appears that the observed risk-adjusted disparities stem primarily from decisions made at the initial hearing, a phenomenon we continue to explore below. Table A4 shows the coefficients for all three models for the outcome of release at the initial hearing, and Tables A5 and A6 show the corresponding coefficients for release after the initial hearing and release at any point, respectively.

AUSA motions for detention

In cases where the AUSA moves for detention, the defendant is ultimately held for detention at the initial hearing 98% of the time. Conversely, in cases where the AUSA does not move for detention, the defendant is released 92% of the time. Given the concordance between the AUSA's actions and the release decision at the initial hearing, disparities in the AUSA's decision are likely an important driver of the observed disparities in risk-adjusted release rates at the initial hearing. Indeed, we find that the AUSA is 4pp (SE = 2pp) more likely to move for detention of Black defendants compared to comparably risky white defendants, with a larger 12pp (SE = 2pp) risk-adjusted disparity for Hispanic defendants. These racial gaps persist across a range of risk levels for both Hispanic and Black defendants, although the disparity between Black and white defendants is largest for low-to-moderate risk scores (Figure 3).

We find that the statutory entitlement to a detention hearing is a strong predictor of the AUSA's decision to move for detention. In cases where the AUSA is entitled to a detention hearing, they move for detention 85% of the time, and, in cases where the AUSA is not entitled, they move for detention 30% of time. After additionally adjusting for the entitlement to a detention hearing, the racial disparities in the rates at which the AUSA moves for detention among similarly risky defendants shrinks, from 12pp (SE = 2pp) to 7pp (SE = 2pp) for Hispanic defendants, and from 4pp (SE = 2pp) to 3pp (SE = 1pp) for Black defendants (Table A7).

It thus appears that the statutory mandate itself is critical for understanding disparities in detention decisions. Regressing the presence of an entitlement to a detention hearing on race/ethnicity and the PTRA criminal history score, we find that Hispanic defendants are 15pp (SE = 2pp) more likely than similarly risky white defendants to be charged with offenses that trigger the entitlement, with a statistically insignificant 2pp disparity for Black defendants (SE = 2pp).

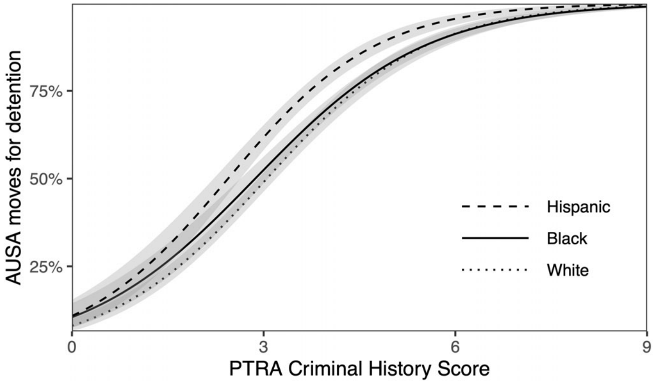

FIGURE 3 Estimated rates at which the assistant US attorney (AUSA) moves for detention at the initial hearing by race/ethnicity and risk score, with 95% confidence bands. Rates are estimated via logistic regression. The AUSA is more likely to move for detention of Black and Hispanic defendants than similarly risky white defendants. After adjusting for the presence of a charge that triggers the AUSA entitlement to a detention hearing, the observed disparities shrink substantially, suggesting that the entitlement policy may itself impose a racial disparate impact (Table A7). Risk-adjusted disparities in motions for detention translate to similar disparities in actual release rates at the initial hearing (Table A4).

The entitlement provision is ostensibly designed to capture some notion of public safety risk (e.g., it applies to offenses deemed "violent"), but the mandate itself seems only loosely connected to statistical risk, and appears to place disproportionate burdens on racial minorities.

Finally, we fit a disparate treatment regression of AUSA motions for detention, adjusting for all observable case features. We find evidence for at most modest effects of (perceptions of) race on decisions, with no statistically significant difference between Black and white defendants, and a statistically significant 3pp (SE = 1.5pp) gap between Hispanic and white defendants. Accordingly, the substantial risk-adjusted disparities we see in AUSA actions appear to stem not from racial animus per se, but rather from policies and practices—including the entitlement provision—that lead to gaps unjustified by concerns for public safety and nonappearance. Table A7 summarizes the above results, showing the coefficients from all four linear probability models when fit to the outcome of the AUSA motion for detention.

PTS recommendations

Like the AUSA, PTS is an important actor in the pretrial process, and the PTS recommendation is a strong predictor of actual detention and release outcomes after the PTS investigation. In cases where PTS recommends detention, the defendant is detained 85% of the time. Conversely, in cases where PTS recommends release, the defendant is released 81% of the time. Much like the AUSA's decision to move for detention and the corresponding release decision at the initial hearing, the PTS recommendation could drive disparities in release rates after the PTS investigation. We find that PTS is 5pp (SE = 2pp) less likely to recommend release for Black defendants compared to similarly risky white defendants, with no evidence of disparate impact for Hispanic defendants (Table A9). However, Black defendants are in reality no less likely to be released after the PTS investigation than similarly risky white defendants, suggesting a corrective mechanism for the observed risk-adjusted disparity in PTS recommendations (Table A10).

We trace the bulk of the risk-adjusted disparity in PTS recommendations for release to the presence of at least one violation of a prior term of release on probation or parole. After additionally adjusting for the presence of a violation, the 5pp risk-adjusted disparity shrinks to a statistically insignificant 2pp (SE = 2pp) (Table A9). Given widespread racial inequities in policing practices, Black defendants on probation or parole may be subject to greater scrutiny than white defendants, magnifying the difference between true violation rates and recorded violation rates among Black defendants. Consistent with this possibility, we indeed find that Black defendants are 12pp more likely to have a prior supervision violation than white defendants of comparable risk (SE = 1pp). Thus, a reliance on prior violations in making detention recommendations (in addition to statistical risk) could lead to unjustified racial disparities in recommendation rates.

Risk-based release policies

One potential strategy to mitigate risk-adjusted disparities in release rates is to presumptively release all defendants below a certain risk score threshold ("threshold policy"). After separately modeling the likelihood of adverse events among released defendants via logistic regression, we can estimate the probability of these adverse events among detained defendants if they had instead been released. In turn, these probabilities allow us to estimate the marginal increase in violation rates that would occur under a threshold policy. Of course, this approach likely suffers from omitted variable bias, as unobserved covariates may explain why one defendant is detained while another is released, even if those two defendants have similar values of observed covariates. We simulate the effects of omitted variables on the likelihood of adverse events by recalculating our estimates after inflating the estimated odds of violation by a constant factor.

Figure 4 shows the results of this analysis. The leftmost "~" label indicates that we observe just under 250 adverse events in our partner partner jurisdiction among the analyzed sample. As the risk-based threshold for release is raised, there is an increase in the estimated number of adverse events. For example, compared to historical practice, we estimate that the policy of releasing all defendants with a PTRA criminal history score less than or equal to four would have led to approximately 750 additional released defendants at a cost of approximately 100 additional adverse events. If we arbitrarily inflate the estimated odds of violation among historically detained defendants by a factor of 10, simulating the impact of an unobserved covariate very strongly correlated with risk, we would expect approximately 350 additional adverse events.

We additionally note the implications of a risk-based threshold policy on observed racial disparities in release rates. White defendants do, on average, have lower risk scores than Black or Hispanic defendants (Figure A2). So, any policy that unilaterally releases lower risk defendants should increase raw release rates of white defendants to a greater extent than that for Black and Hispanic defendants. However, so long as white defendants are released at the same or higher rates than Black or Hispanic defendants across all risk scores, which we approximately observe (Figure A3), a unilateral threshold release policy can only reduce overall risk-adjusted disparities or leave them unchanged. That being said, eliminating disparities with respect to one measure is unlikely to eliminate disparities with respect to another. For example, while a unilateral threshold policy based on PTRA scores may reduce disparities conditional on

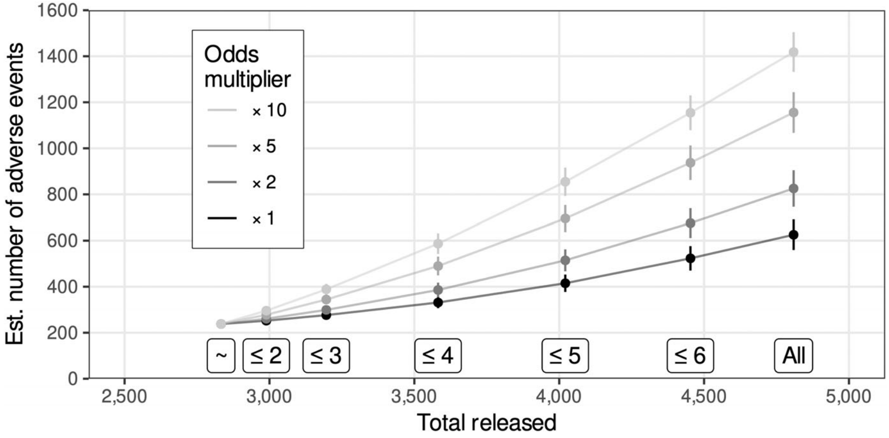

FIGURE 4 Estimated number of adverse events (defined as a bond revocation, failure to appear, or rearrest) under hypothetical risk-based release policies, with 95% normal confidence intervals for the estimated number of violations. Estimates are generated via a logistic regression model trained on an extensive set of risk-relevant covariates derived from prior criminal history and case characteristics among released defendants. Standard errors are derived from posterior simulations of the logistic regression coefficients via the sim function from the arm R package (Gelman & Su, 2021). To account for potential omitted variable bias in the model used to estimate the likelihood of violating, the plot also shows the estimated number of adverse events after multiplying each newly released defendant's estimated odds of violation by a constant factor. "∼" refers to the historical release practice of the partner jurisdiction among the analysis sample. "All" refers to the hypothetical policy of releasing all defendants. Each digit *X* refers to the policy of releasing all defendants with a Federal Pretrial Risk Assessment (PTRA) criminal history score of *X* or lower, in addition to the defendants who were actually released by the partner jurisdiction. Policies for PTRA criminal history scores of one, seven, and eight are omitted due to similarity with the terminal policies.

PTRA scores, it could potentially exacerbate disparities conditional on a different estimate of risk.

Our analysis is designed to estimate the costs and benefits of different threshold policies. It shows that a threshold policy would allow for the release of a significant number of defendants without increasing the projected societal harm substantially. That said, determining the appropriate threshold is ultimately a policy question, and thus cannot be answered here conclusively.

DISCUSSION

Our results offer several potential interventions through which the pretrial process could be improved by reducing disparities while reducing (or not substantially increasing) the risk to public safety. First, the statutory mandate that lowers the procedural bar for AUSAs to move for detention in certain types of cases could be reconsidered. The case-specific carve-outs appear to be only loosely connected to pretrial risk, but disproportionately involve Black and Hispanic defendants. Hence, they can exacerbate racial disparities without reducing adverse pretrial outcomes. Second, it is important to identify and make salient by way of empirical research the racial gaps at the initial hearing. Awareness of these disparities is a necessary requirement for all officers of the court—including prosecutors, judges, and defense attorneys—to adjust and mitigate the harm caused by unjustified disparities. For instance, in light of the statistical evidence, it appears particularly important to ensure detention at the initial hearing is predicated on clearly articulated reasons justified by the particularities of a case. Among others, judges could require that they and AUSAs be provided with assessments of pretrial risk at the initial hearing. The US attorney's office could take steps to better understand and address the source of the unexplained gap in motions for detention, for example, by thoroughly recording the outcomes and case characteristics of critical pretrial decisions and periodically auditing office practices to determine if there are unjustified disparate impacts on certain groups of defendants. Third, although risk-adjusted disparities in PTS recommendations do not appear to materialize in judicial decisions after the PTS investigation, it appears necessary to identify and address the potential sources of those disparities, such as prior violations of release on probation or parole. PTS could, for example, adopt a more structured decision-making process that better aligns recommendations with observed risk.

There are several key limitations to our analysis. First, PTRA scores do not fully capture all risk-relevant information that is observable at the individual decision points. For example, gang affiliation does not directly factor into the calculation of PTRA scores, and gang affiliation could ostensibly increase public safety risk. Although a judge may incorporate knowledge of a gang affiliation into her decision-making process for a particular defendant, the defendant's PTRA score would not reflect this knowledge. As an initial robustness check to this concern, we find that our results are qualitatively similar across three different risk scores that account for different sets of observed covariates (Tables A12–A18).

We further assess the robustness of our approach to the inclusion of our own, estimated risk scores. To that end, we train a cross-validated ridge regression model that predicts the likelihood of adverse outcomes among released defendants using the comprehensive set of all observed, risk-relevant covariates, including specific charge identifiers, prior criminal history, and the presence of pending charges, warrants, or detainers. After adjusting for the model-predicted likelihood of adverse outcomes among all defendants, estimates of disparate impact are even larger in magnitude than the estimates in the main analysis (see Tables A12–A18). Although these tests demonstrate the robustness of our findings to specifications relying on observed characteristics, we nonetheless cannot rule out with certainty that covariates unobserved by us, though observable to decision-makers, could feasibly explain at least part of the observed risk-adjusted disparities. We note, however, that an extensive body of prior research indicates that risk assessment tools are better than unaided experts at predicting a range of outcomes, including recidivism (c.f. Goel et al., 2021 for a review of the literature), making it unlikely that the disparities we observe are fully explained by judges having access to additional risk-relevant information.

Second, while PTRA risk scores estimate nonappearance and safety risk, they do not provide an estimate of the severity of a potential nonappearance violation or reoffense. For example, while two defendants may have the same PTRA score and therefore the same estimated likelihood of reoffending, one defendant may be more likely to commit a felony offense than the other. Officers of the court may be able to accurately distinguish between these two risk categories, which could help explain observed risk-adjusted disparities. It should be noted, however, that our results are robust to adjustment for Public Safety Assessment (PSA) scores which, in addition to estimating the failure to appear and to commit new crimes, separately estimate the risk of new violent criminal activity.

Third, calculating risk scores can be a context-dependent undertaking, and information gleaned from matched bail reports may not always provide enough contextual information to match the PTS determination. For example, multiple counts of a conviction resulting from a single arrest date should only count as a single conviction in the calculation of PTRA risk scores. Although we can verify that estimation error for PTRA scores is small for the subset of defendants for whom PTS has calculated true PTRA scores, it is impossible for us to verify the PTRA calculation for defendants without true scores.

Fourth, risk scores trained on historical violations may be biased if the proportion of outcomes that are mislabeled differs across groups. For example, if Black and Hispanic defendants are more likely than white defendants to be caught for the same technical violation or rearrested for the same offense, it is possible that lower risk scores for white defendants may not actually be indicative of lower risk.

Fifth, our model for predicting violations may suffer from omitted variable bias, since released defendants may systematically differ from detained defendants in a way that we cannot observe. If violation rates are influenced by at least one unobserved covariate that differs in prevalence among released and detained defendants, then our estimates of violation rates among detained defendants would be biased.

Finally, while we can estimate the number of additional violations that would occur under hypothetical risk-based threshold release policies, it appears at least possible that some of the defendants that are released under the current policy would instead be detained under a threshold policy. For example, if the court implemented a threshold policy and released all defendants with PTRA scores below a particular threshold, it could simultaneously decide to release fewer (or more) defendants above the threshold than is true under the current policy. Our estimates are thus based on the assumption that the court does not change its release practice for defendants above the threshold.

Overall, despite these limitations, we believe our analysis illustrates the value of taking a holistic view of decision making and discrimination, and we hope our approach provides a roadmap for investigating disparities throughout the criminal justice system and beyond.

ACKNOWLEDGMENTS

The authors thank Amit Haim for valuable research assistance. We also thank the anonymous district court judges, magistrate judges, PTS officers, IT staff, and court experts from our partner jurisdiction for their extensive feedback and guidance. Finally, we thank our discussant David Bjerk and audience members at the 15th Conference on Empirical Legal Studies for their helpful feedback and comments.

REFERENCES

Abrams, D. S., Bertrand, M., & Mullainathan, S. (2012). Do judges vary in their treatment of race? *The Journal of Legal Studies*, *41*(2), 347–383.

Advancing Pretrial Policy & Research. (n.d.). *About the Public Safety Assessment*. https://advancingpretrial.org/psa/factors/

Antonovics, K., & Knight, B. G. (2009). A new look at racial profiling: Evidence from the Boston Police Department. *The Review of Economics and Statistics*, *91*(1), 163–177.

Anwar, S., Bayer, P., & Hjalmarsson, R. (2012). The impact of jury race in criminal trials. *The Quarterly Journal of Economics*, *127*(2), 1017–1055.

Arnold, D., Dobbie, W., & Hull, P. (2021). Measuring racial discrimination in algorithms. In *American Economic Association papers and proceedings* (Vol. 111, pp. 49–54). https://www.aeaweb.org/articles?id=10.1257/pandp.20211080

Arnold, D., Dobbie, W., & Hull, P. (2022). Measuring racial discrimination in bail decisions. *American Economic Review*, *112*(9), 2992–3038.

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics*, *133*(4), 1885–1932.

Austin, A. (2017). The presumption for detention statute's relationship to release rates. *Federal Probation*, *81*, 52.

Ayres, I. (2005). Three tests for measuring unjustified disparate impacts in organ transplantation: The problem of "included variable" bias. *Perspectives in Biology and Medicine*, *48*(1), 68–S87.

Ayres, I., & Waldfogel, J. (1993). A market test for race discrimination in bail setting. *Stanford Law Review*, *46*, 987.

Bohren, J. A., Hull, P., & Imas, A. (2022). *Systemic discrimination: Theory and measurement*. National Bureau of Economic Research (NBER) Working Paper. https://www.nber.org/papers/w29820

Browne, G. E., & Strong, S. M. (2022). *Pretrial release and misconduct in federal district courts, fiscal years 2011–2018*. Bureau of Justice Statistics (BJS). https://bjs.ojp.gov/library/publications/pretrial-release-and-misconduct-federal-district-courts-fiscal-years-2011-2018

Cadigan, T. P., Johnson, J. L., & Lowenkamp, C. T. (2012). The re-validation of the Federal Pretrial services Risk Assessment (PTRA). *Federal Probation*, *76*, 3.

Carr, J. G. (2016). Why pretrial release really matters. *Federal Sentencing Reporter*, *29*, 217.

Chohlas-Wood, A., Nudell, J., Yao, K., Lin, Z., Nyarko, J., & Goel, S. (2021). Blind justice: Algorithmically masking race in charging decisions. In *Proceedings of the 2021 aaailacm conference on ai, ethics, and society* (pp. 35–45). ACM (Association for Computing Machinery).

Corbett-Davies, S., & Goel, S. (2018). The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*.

DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., Comfort, M., & Misra, S. (2020). *The public safety assessment: A re-validation and assessment of predictive utility and differential prediction by race and gender in Kentucky* (Vol. 19, pp. 409–431). Criminology and Public Policy.

Demuth, S. (2003). Racial and ethnic differences in pretrial release decisions and outcomes: A comparison of Hispanic, Black, and White felony arrestees. *Criminology*, *41*(3), 873–908.

Demuth, S., & Steffensmeier, D. (2004). Ethnicity effects on sentence outcomes in large urban courts: Comparisons among White, Black, and Hispanic defendants. *Social Science Quarterly*, *85*(4), 994–1011.

Didwania, S. H. (2020). The immediate consequences of federal pretrial detention. *American Law and Economics Review*, *22*(1), 24–74.

Ding, P., & Miratrix, L. W. (2015). To adjust or not to adjust? Sensitivity analysis of M-bias and butterfly-bias. *Journal of Causal Inference*, *3*(1), 41–57.

Dobbie, W., Goldin, J., & Yang, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, *108*(2), 201–240.

Fryer, R. G., Jr. (2019). An empirical analysis of racial differences in police use of force. *Journal of Political Economy*, *127*(3), 1210–1261.

Gaebler, J., Cai, W., Basse, G., Shroff, R., Goel, S., & Hill, J. (2022). A causal framework for observational studies of discrimination. In *Statistics and public policy* (Vol. 9). Taylor and Francis Online.

Gelman, A., & Su, Y.-S. (2021). arm: Data analysis using regression and multilevel/hierarchical models [Computer software manual] (R package version 1.12-2). https://CRAN.R-project.org/package=arm

Goel, S., Shroff, R., Skeem, J., & Slobogin, C. (2021). The accuracy, equity, and jurisprudence of criminal risk assessment. In *Research handbook on big data law*. Edward Elgar Publishing.

Grau, N., Vergara, D. (2020). *A simple test for prejudice in decision processes: The prediction-based outcome test*. Working Papers wp493, University of Chile, Department of Economics. https://ideas.repec.org/p/udc/wpaper/wp493.html

Greiner, D. J., & Rubin, D. B. (2011). Causal effects of perceived immutable characteristics. *Review of Economics and Statistics*, *93*(3), 775–785.

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

Hull, P. (2021). *What marginal outcome tests can tell us about racially biased decision-making*. National Bureau of Economic Research (NBER) Working Paper. https://www.nber.org/papers/w28503

Jung, J., Corbett-Davies, S., Shroff, R., & Goel, S. (2019). Omitted and included variable bias in tests for disparate impact. *arXiv preprint arXiv:1809.05651*.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, *133*(1), 237–293.

Kurlychek, M. C., & Johnson, B. D. (2019). Cumulative disadvantage in the American criminal justice system. *Annual Review of Criminology*, *2*, 291–319.

Kutateladze, B., Lynn, V., & Liang, E. (2012). *Do race and ethnicity matter in prosecution? A review of empirical studies*. Vera Institute of Justice.

Kutateladze, B. L., Andiloro, N. R., Johnson, B. D., & Spohn, C. C. (2014). Cumulative disadvantage: Examining racial and ethnic disparity in prosecution and sentencing. *Criminology*, *52*(3), 514–551.

Lakkaraju, H., Kleinberg, J., Leskovec, J., Ludwig, J., & Mullainathan, S. (2017). The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 275–284). Association for Computing Machinery (ACM).

Laura and John Arnold Foundation. (2013). Developing a national model for pretrial risk assessment. https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/LJAF-research-summary_PSA-Court_4_1.pdf

Leslie, E., & Pope, N. G. (2017). The unintended impact of pretrial detention on case outcomes: Evidence from NYC arraignments. *The Journal of Law and Economics*, *60*, 529–557.

Lin, Z., Jung, J., Goel, S., & Skeem, J. (2020). The limits of human predictions of recidivism. *Science Advances*, *6*(7), eaaz0652.

Lowenkamp, C. T. (2009). The development of an actuarial risk assessment instrument for US Pretrial Services. *Federal Probation*, *73*, 33.

Lowenkamp, C. T., VanNostrand, M., & Holsinger, A. M. (2013). *The hidden costs of pretrial detention*. LJAF.

Lynch, M. J., & Patterson, E. B. (1991). *Race and criminal justice*. Harrow and Heston New York.

Metcalfe, C., & Chiricos, T. (2018). Race, plea, and charge reduction: An assessment of racial disparities in the plea process. *Justice Quarterly*, *35*(2), 223–253.

Oleson, J. C., Lowenkamp, C. T., Cadigan, T. P., VanNostrand, M., & Wooldredge, J. (2016). The effect of pretrial detention on sentencing in two federal districts. *Justice Quarterly*, *33*(6), 1103–1122.

Omori, M., & Johnson, O. (2019). Racial inequality in punishment. In *Oxford research encyclopedia of criminology and criminal justice*. Oxford University Press.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, *4*(7), 736–745.

Rehavi, M. M., & Starr, S. B. (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy*, *122*(6), 1320–1354.

Schlesinger, T. (2005). Racial and ethnic disparity in pretrial criminal processing. *Justice Quarterly*, *22*(2), 170–192.

Schlesinger, T. (2007). The cumulative effects of racial disparities in criminal processing. *The Journal of the Institute of Justice & International Studies*, *7*, 261.

Spohn, C. (2008). Race, sex, and pretrial detention in federal court: Indirect effects and cumulative disadvantage. *University of Kansas Law Review*, *57*, 879.

Stevenson, M. T. (2018). Distortion of justice: How the inability to pay bail affects case outcomes. *The Journal of Law, Economics, and Organization*, *34*(4), 511–542.

Stolzenberg, L., D'Alessio, S. J., & Eitle, D. (2013). Race and cumulative discrimination in the prosecution of criminal defendants. *Race and Justice*, *3*(4), 275–299.

Sutton, J. R. (2013). Structural bias in the sentencing of felony defendants. *Social Science Research*, *42*(5), 1207–1221.

Turner, K., & Johnson, J. B. (2007). The relationship between type of attorney and bail amount set for Hispanic defendants. *Hispanic Journal of Behavioral Sciences*, *29*(3), 384–400.

How to cite this article: Grossman, J., Nyarko, J., & Goel, S. (2023). Racial bias as a multi-stage, multi-actor problem: An analysis of pretrial detention. *Journal of Empirical Legal Studies*, 1–48. https://doi.org/10.1111/jels.12343

APPENDIX A

Figure A1 is an expanded version of Figure 1 that shows all major pretrial decision points. Figure A2 shows the distribution of PTRA criminal history scores by race/ethnicity. Figure A3 shows estimated overall release rates by race/ethnicity and PTRA criminal history score. Figures A4–A6 highlight the similar performance of PTRA scores, PTRA criminal history scores, and PSA scores in predicting adverse events in our partner jurisdiction. Figure A7 shows that risk scores are calibrated across race and ethnicity groups.

Table A1 shows summary statistics for the released and detained defendants in the partner jurisdiction. Tables A2 and A3 show the coefficients of the cross-validated L2 (ridge) regression model used to estimate in-sample risk of adverse events among the released study population. Tables A4–A10 show the coefficients from the main disparate impact and disparate treatment analyses as referenced in the text, with heteroskedasticity-robust standard errors. When judge fixed effects are included in the regression, standard errors are clustered by judge. Table A11 shows that our main disparate impact results are robust to the inclusion of higher order terms for risk.

The main disparate treatment analyses already adjust for PTRA scores, PTRA criminal history scores, and PSA scores. Thus, Tables A12–A18 compare the results of only the disparate impact analyses across different risk scores. For comparison, we also include an adjustment for the estimated probability that a given defendant either fails to appear, reoffends, or has bond revoked if released. This probability is estimated via a cross-validated ridge regression model trained on a comprehensive set of risk-relevant covariates observed for all released defendants, such as specific charge identifiers, prior criminal history, and pending charges, warrants, and detainers. The results are qualitatively similar across risk scores.

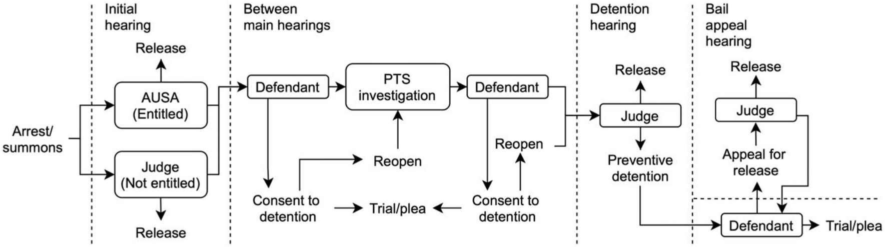

FIGURE A1 Expanded version of Figure 1. Additions include (1) consenting to detention after the pretrial services (PTS) investigation, (2) reopening bail proceedings after defendant consents to detention, and (3) appeals for release following the detention hearing. Our data only shows "successful" reopenings of bail proceedings following consent after the PTS investigation and "successful" appeals for release. For example, if bail proceedings were reopened after the defendant consented to detention, but the defendant was detained at the resulting detention hearing, we would not see that a detention hearing was held. However, if the same defendant was released at the detention hearing, we would see a record of the detention hearing. Thus, while we cannot reliably investigate individual decision points after the investigation, we can still reliably model the more general outcome of release after the PTS investigation.

TABLE A1 Summary statistics by release status for the 4809 cases in our main analysis

	Detained	Released
Total	1977	2832
Female	7%	22%
Black	44%	29%
Hispanic	30%	22%
White	26%	49%
Age	34	39
PTRA score	8.1	5.3
PTRA criminal history scores	5.0	3.0
PSA NCA score	4.9	2.1
PSA FTA score	2.0	0.9
PSA NVCA score	2.6	1.1
Prior felonies	3.4	1.0
Prior misdemeanors	3.2	1.5
Prior FTAs	2.0	0.7
Has prior violation	74%	27%
Has pending charge	27%	11%
Has prior sentence >2 weeks	80%	33%
Has felony charge	99%	84%
Has firearms charge	31%	11%
Has drug charge	33%	32%
Released at initial hearing	0%	61%
Has detention hearing	77%	39%
AUSA entitled	81%	47%
AUSA moves for detention	99%	37%
Presumption for detention	42%	35%
Any violation (including technical)	0%	28%
Bond revoked	0%	6%
FTA rate	0%	2%
Rearrest rate	0%	2%
Bond revocation, FTA, or rearrest	0%	8%

Note: Means are shown for continuous variables, and proportions for binary variables.
Abbreviations: AUSA, assistant US attorney; FTA, failure to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

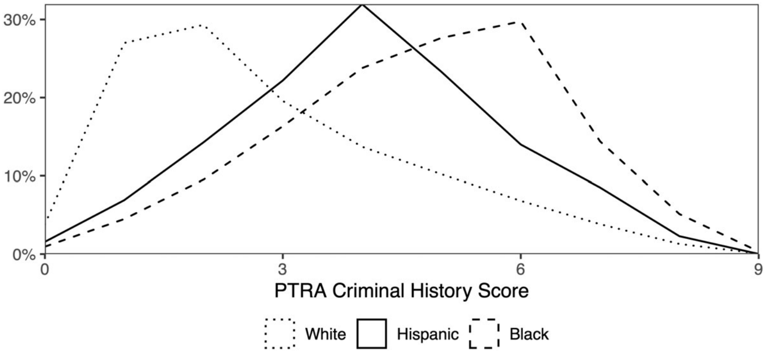

FIGURE A2 Observed frequency of PTRA criminal history risk scores across Black, Hispanic, and white defendants. White defendants have, on average, the lowest risk scores, and Hispanic defendants have lower average risk scores than Black defendants. PTRA, Federal Pretrial Risk Assessment

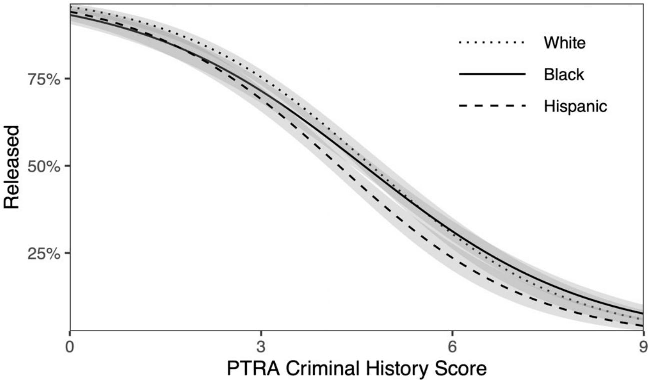

FIGURE A3 Release rates by PTRA criminal history score and race/ethnicity, estimated via logistic regression. Adjusting for risk scores, Hispanic defendants are released at lower rates than white defendants, with the most pronounced difference for risk scores above 3. For risk scores below 4, Black defendants also appear to be released at lower rates than similarly risky white defendants, though the cumulative risk-adjusted release rates for Black and white defendants are statistically indistinguishable. PTRA, Federal Pretrial Risk Assessment

TABLE A2 First subset of coefficient estimates from the L2 (ridge) regression model trained on released defendants using the outcome of any bond revocation, failure to appear, or new arrest.

Term	Coefficient
(Intercept)	−2.703
Black	0.030
Hispanic	−0.002
Age at activation	−0.001
Male	0.009
PTRA	0.015
Criminal history score	0.021
PSA FTA	0.031
PSA NCA	0.014
PSA NVCA	0.026
Has detainer	0.075
Has warrant	0.079
Has pending charge	0.069
Has prior sentence geq 14 Days	0.052
Has violation	0.060
One FTA within 2 years	0.122
Two FTAs within 2 years	0.142
More than two FTAs within 2 years	0.096
One FEL conviction	0.019
Two FEL convictions	0.028
Three FEL convictions	0.033
Four FEL convictions	0.052
More than four FEL convictions	0.069
One MSD conviction	−0.011
Two MSD convictions	0.021
Three MSD convictions	0.049
Four MSD convictions	0.092
More than four MSD convictions	0.076

Note: The second subset of coefficients is in Table A3.
Abbreviations: FEL, felony; FTA, failure to appear; MSD, misdemeanor; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

TABLE A3 Second subset of coefficient estimates from the L2 (ridge) regression model trained on released defendants using the outcome of any bond revocation, failure to appear, or new arrest.

Term	Coefficient
Fel Intx	–0.059
Fel Sex	0.018
Fel Meth	0.037
Fel Frau	–0.011
Fel Robb	0.111
Fel Maif	–0.023
Fel Otdr	–0.028
Fel Misc	–0.023
Mis Traf	–0.059
Mis Misc	–0.042
Fel Fire	0.063
Fel Mait	0.000
Fel Assa	0.096
Fel Immi	–0.065
Fel Heri	0.012
Fel Mari	–0.025
Fel Larc	0.042
Fel Coca	0.020
Fel Opia	0.018
Fel Coun	–0.013
Fel Rack	–0.024
Fel Homi	0.145
Fel Embz	–0.054
Fel Other	0.057
Mis Larc	–0.018
Mis Burg	0.016
Mis Assu	–0.061
Mis Frau	–0.056
Mis Assa	0.021
Mis Other	–0.037

Note: The first subset of coefficients is in Table A2.

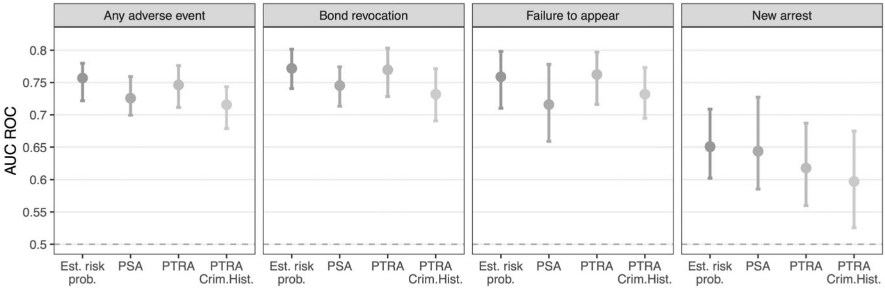

FIGURE A4 Performance of estimated risk probabilities, imputed PSA FTA/NCA/NVCA scores, PTRA scores, and PTRA criminal history scores with respect to predicting bond revocations, FTAs, and rearrests of released defendants in our partner jurisdiction, with 95% normal confidence intervals. No risk score performs significantly better than another. The models perform best for predicting bond revocations and failures to appear, but do not perform nearly as well for predicting new arrests. AUC ROC is the probability of correctly classifying a randomly drawn defendant with a violation as riskier than a randomly drawn defendant without a violation. Higher AUC ROC values imply better performance with respect to predicting violations. An AUC ROC of 0.5 indicates that the classifier is no better than random guessing. AUC ROC values are computed via logistic regression of the adverse outcome on the given risk score(s). Standard errors are calculated from 100 bootstrapped samples of released defendants. AUC ROC, area under the receiver operating characteristic curve; FTA, failures to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

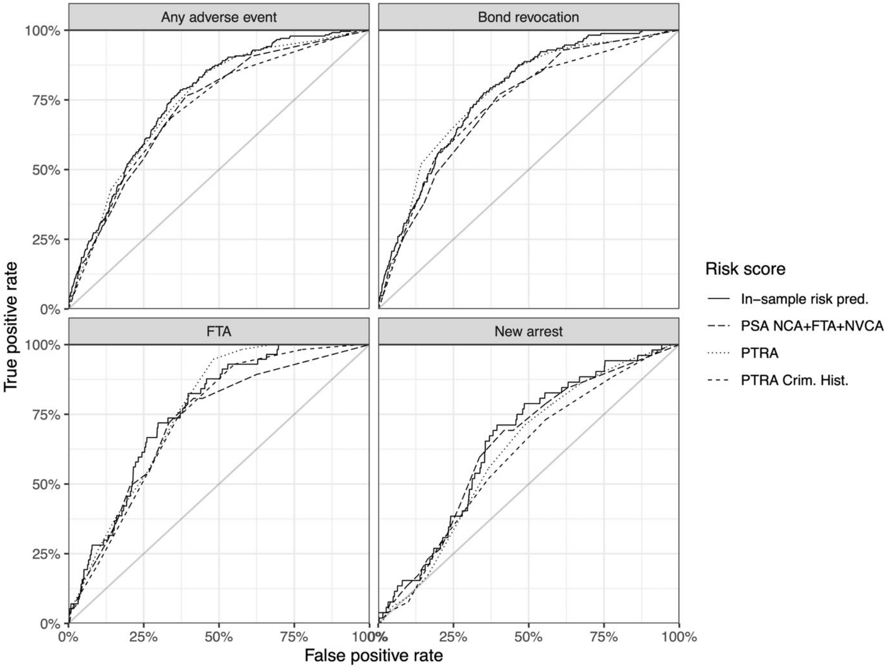

FIGURE A5 ROC curves of estimated risk probabilities, imputed PSA FTA/NCA/NVCA scores, PTRA scores, and PTRA criminal history scores with respect to predicting bond revocations, FTAs, and rearrests of released defendants in our partner jurisdiction. For each risk score, the curve is generated by examining all possible risk score thresholds for classification of each defendant as either violating or not violating. ROC curves closer to the top left are generally more desirable, but the precise tradeoff between the false positive rate and true positive rate is decided by the policymaker. The curves indicate that the PTRA and PSA risk scores perform approximately as well as a risk model trained on the same sample of released defendants that the PTRA and PSA risk scores are evaluated on. FTA, failures to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment; ROC, receiver operating characteristics.

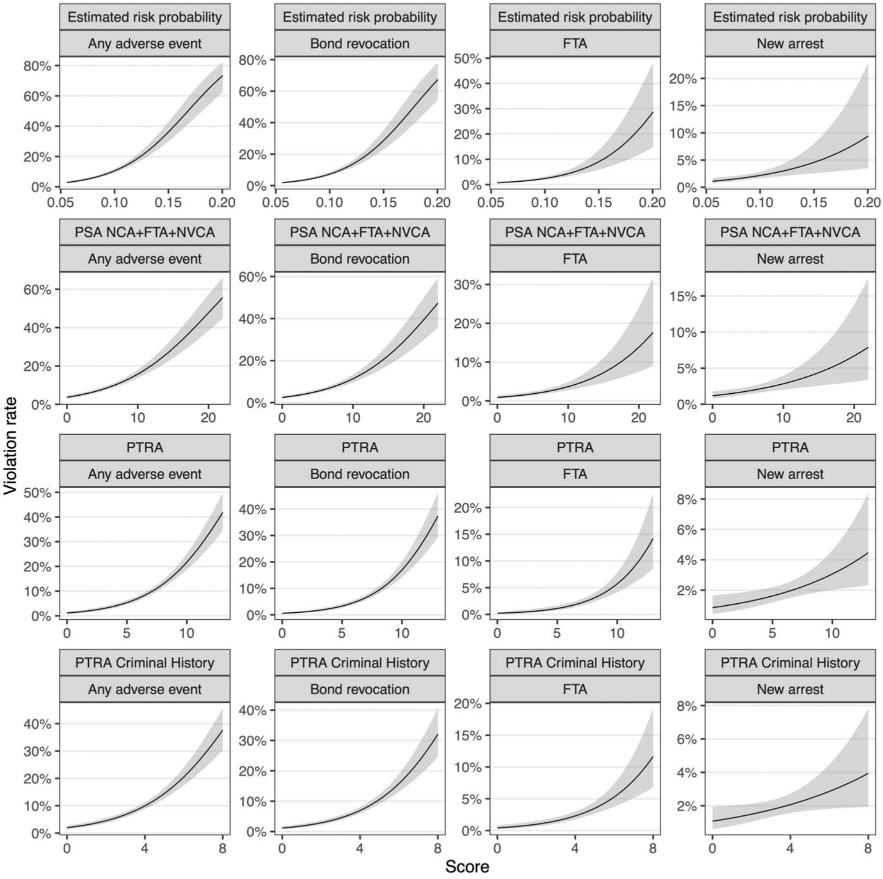

FIGURE A6 Logistic regressions of bond revocation, FTA, new arrest, or any adverse event on in-sample estimated risk probabilities, imputed PSA FTA/NCA/NVCA scores, PTRA scores, and PTRA criminal history scores. Each risk score correlates positively with each violation type. FTA, failures to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

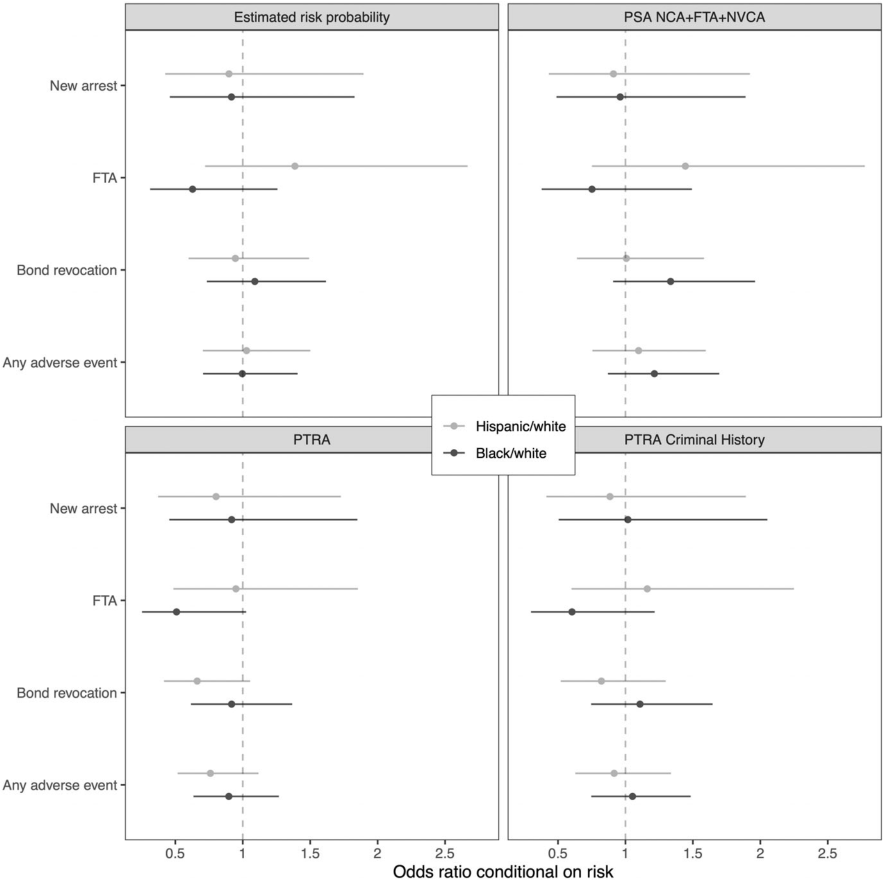

FIGURE A7 Ratio of odds of bond revocation, FTA, and new arrest for Hispanic versus white defendants and Black versus white defendants. Odds are derived from logistic regression of the each violation on each risk score and race. None of the odds ratios are significantly different from zero at a 95% confidence level, indicating that the risk scores are calibrated across race groups. Although the PSA and full PTRA scores are borderline calibrated for the bond revocation and FTA outcomes, respectively, the PTRA criminal history scores are well within the margin of insignificance. This further supports our use of PTRA criminal history scores in the main analysis. FTA, failures to appear; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

TABLE A4 For the outcome of release at the initial hearing, coefficients from the main disparate impact and disparate treatment analyses. Coefficients are accompanied by heteroskedasticity-robust standard errors.

	Outcome: Released at initial hearing		
	Unadjusted **(1)**	**Disparate impact** **(2)**	**Disparate treatment** **(3)**
Hispanic–white disparity	−0.297***	−0.132***	−0.035*
	(0.017)	(0.015)	(0.014)
Black–white disparity	−0.286***	−0.027	0.008
	(0.016)	(0.015)	(0.014)
PTRA criminal history score		−0.135***	0.029**
		(0.003)	(0.009)
Constant	0.532***	0.916***	0.414***
	(0.011)	(0.013)	(0.115)
Month and year fixed effects			X
Judge and city fixed effects			X
Additional risk scores			X
Age and sex controls			X
Criminal history controls			X
Charge fixed effects			X
Observations	4809	4809	4809
R^2	0.088	0.333	0.548
Adjusted R^2	0.088	0.333	0.539
Residual std. error	0.458	0.391	0.325

Note: Adjusting for Federal Pretrial Risk Assessment (PTRA) criminal history scores, we find that Hispanic defendants are 13pp less likely to be released at the initial hearing than similarly risky white defendants (SE = 2pp), with a directional 3pp gap for Black defendants (SE = 2pp, $p = 0.06$). To estimate disparate treatment, we adjust for a full set of controls, finding than Hispanic defendants are released less often than similarly situated white defendants, with statistically insignificant estimated effects for Black defendants.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

TABLE A5 For the outcome of release after the initial hearing, coefficients from the main disparate impact and disparate treatment analyses. Coefficients are accompanied by heteroskedasticity-robust standard errors.

	Outcome: Released after initial hearing		
	Unadjusted **(1)**	**Disparate impact** **(2)**	**Disparate treatment** **(3)**
Hispanic–white disparity	−0.048*	0.002	−0.028
	(0.023)	(0.023)	(0.024)
Black–white disparity	−0.116***	−0.014	0.007
	(0.021)	(0.022)	(0.022)
PTRA criminal history score		−0.077***	−0.100***
		(0.005)	(0.014)
Constant	0.429***	0.728***	0.458**
	(0.017)	(0.027)	(0.172)
Month and year fixed effects			X
Judge and city fixed effects			X
Additional risk scores			X
Age and sex controls			X
Criminal history controls			X
Charge fixed effects			X
Observations	3042	3042	3042
R^2	0.010	0.072	0.257
Adjusted R^2	0.009	0.071	0.237
Residual std. error	0.480	0.464	0.421

Note: Unlike the corresponding analyses for release at the initial hearing, we do not obtain statistically significant estimates of disparate treatment or disparate impact for release after the initial hearing.
Abbreviation: PTRA, Federal Pretrial Risk Assessment.
$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

TABLE A6 For the outcome of release at any point in the pretrial process, coefficients from the main disparate impact and disparate treatment analyses. Coefficients are accompanied by heteroskedasticity-robust standard errors.

	Outcome: Released at any point		
	Unadjusted **(1)**	**Disparate impact** **(2)**	**Disparate treatment** **(3)**
Hispanic–white disparity	−0.209***	−0.059***	−0.038*
	(0.018)	(0.017)	(0.017)
Black–white disparity	−0.250***	−0.016	0.014
	(0.016)	(0.016)	(0.016)
PTRA criminal history score		−0.122***	−0.070***
		(0.003)	(0.011)
Constant	0.729***	1.080***	0.678***
	(0.010)	(0.011)	(0.127)
Month and year fixed effects			X
Judge and city fixed effects			X
Additional risk scores			X
Age and sex controls			X
Criminal history controls			X
Charge fixed effects			X
Observations	4809	4809	4809
R^2	0.054	0.245	0.410
Adjusted R^2	0.054	0.244	0.399
Residual std. error	0.479	0.428	0.382

Note: Overall, Hispanic defendants are 6pp less likely than similarly risky white defendants to be released at any point in the pretrial process, with a statistically insignificant disparity for Black defendants. When adjusting for "kitchen sink" controls, we do not obtain statistically significant estimates of disparate treatment in overall release rates for either Black or Hispanic defendants.

Abbreviation: PTRA, Federal Pretrial Risk Assessment.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE A7 For the outcome of whether or not the AUSA moves for detention, coefficients from the main disparate impact and disparate treatment analyses. Coefficients are accompanied by heteroskedasticity-robust standard errors.

	Outcome: AUSA moves for detention			
	Unadjusted	**Disparate impact**		**Disparate treatment**
	(1)	**(2)**	**(3)**	**(4)**
Hispanic–white disparity	0.282***	0.120***	0.068***	0.031*
	(0.017)	(0.016)	(0.015)	(0.015)
Black–white disparity	0.295***	0.042**	0.034*	−0.003
	(0.016)	(0.015)	(0.014)	(0.014)
PTRA criminal history score		0.132***	0.086***	−0.025**
		(0.003)	(0.004)	(0.010)
AUSA entitled			0.357***	0.158***
			(0.016)	(0.034)
Constant	0.451***	0.074***	0.050***	0.491***
	(0.011)	(0.013)	(0.011)	(0.121)
Month and year fixed effects				X
Judge and city fixed effects				X
Additional risk scores				X
Age and sex controls				X
Criminal history controls				X
Charge fixed effects				X
Observations	4809	4809	4809	4809
R^2	0.086	0.316	0.407	0.511
Adjusted R^2	0.085	0.315	0.407	0.501
Residual std. error	0.463	0.401	0.373	0.342

Note: The AUSA is 12pp more likely to move for detention of Hispanic defendants compared to similarly risky white defendants (SE = 2pp), with a 4pp gap for Black defendants (SE = 2pp). These disparate impacts shrink substantially after adjusting for the presence of a charge that triggers the AUSA's statutory entitlement to a detention hearing, suggesting that the entitlement itself may impose a disparate impact. The risk-adjusted disparities in the AUSA's decision to move for detention translate to similar disparities in the rates at which defendants are released at the initial hearing (Table A4). Furthermore, we find that the AUSA moves for detention of Hispanic defendants more often than similarly situated white defendants, with a statistically insignificant estimated disparate treatment effect for Black defendants.

Abbreviations: AUSA, assistant US attorney; PTRA, Federal Pretrial Risk Assessment.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE A8 For the outcome of consent at any point before the PTS investigation, coefficients from the main disparate impact and disparate treatment analyses. Coefficients are accompanied by heteroskedasticity-robust standard errors.

	Outcome: Consent before investigation		
	Unadjusted **(1)**	**Disparate impact** **(2)**	**Disparate treatment** **(3)**
Hispanic–white disparity	−0.018	−0.011	0.016
	(0.017)	(0.017)	(0.019)
Black–white disparity	−0.037*	−0.024	−0.012
	(0.015)	(0.016)	(0.017)
PTRA criminal history score		−0.010*	0.067***
		(0.004)	(0.010)
Constant	0.157***	0.195***	0.423**
	(0.012)	(0.020)	(0.136)
Month and year fixed effects			X
Judge and city fixed effects			X
Additional risk scores			X
Age and sex controls			X
Criminal history controls			X
Charge fixed effects			X
Observations	3042	3042	3042
R^2	0.002	0.004	0.157
Adjusted R^2	0.001	0.003	0.135
Residual std. error	0.343	0.342	0.319

Note: We do not obtain statistically significant estimates of disparate treatment or disparate impact for consenting to detention before the PTS investigation.

Abbreviation: PTRA, Federal Pretrial Risk Assessment; PTS, Federal Pretrial Services.

$*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

TABLE A9 For the outcome of whether or not PTS recommends release, coefficients from the main disparate impact and disparate treatment analyses. Coefficients are accompanied by heteroskedasticity-robust standard errors.

	Outcome: Pretrial services recommends release			
	Unadjusted	**Disparate impact**		**Disparate treatment**
	(1)	**(2)**	**(3)**	**(4)**
Hispanic–white disparity	−0.066**	−0.008	−0.011	−0.035
	(0.025)	(0.025)	(0.024)	(0.026)
Black–white disparity	−0.178***	−0.054*	−0.017	−0.016
	(0.023)	(0.024)	(0.023)	(0.024)
PTRA criminal history score		−0.091***	−0.054***	−0.079***
		(0.006)	(0.006)	(0.016)
Has prior violation			−0.266***	−0.117***
			(0.023)	(0.028)
Constant	0.501***	0.859***	0.835***	0.627**
	(0.018)	(0.028)	(0.028)	(0.192)
Month and year fixed effects				X
Judge and city fixed effects				X
Additional risk scores				X
Age and sex controls				X
Criminal history controls				X
Charge fixed effects				X
Observations	2654	2654	2654	2654
R^2	0.024	0.108	0.157	0.268
Adjusted R^2	0.023	0.107	0.155	0.245
Residual std. error	0.486	0.464	0.452	0.427

Note: PTS is 5pp less likely to recommend release for Black defendants than white defendants with similar risk scores (SE = 2pp). We do not obtain statistically significant estimates of disparate impact for Hispanic defendants, or disparate treatment of either Black or Hispanic defendants.

Abbreviations: PTRA, Federal Pretrial Risk Assessment; PTS, pretrial services.

*p < 0.05; **p < 0.01; ***p < 0.001.

TABLE A10 For the outcome of release at any point after the PTS investigation, coefficients from the main disparate impact and disparate treatment analyses. Coefficients are accompanied by heteroskedasticity-robust standard errors.

	Outcome: Released after investigation		
	Unadjusted **(1)**	**Disparate impact** **(2)**	**Disparate treatment** **(3)**
Hispanic–white disparity	−0.067**	−0.009	−0.029
	(0.025)	(0.025)	(0.026)
Black–white disparity	−0.148***	−0.023	0.007
	(0.023)	(0.024)	(0.024)
PTRA criminal history score		−0.092***	−0.076***
		(0.006)	(0.015)
Constant	0.500***	0.860***	0.568**
	(0.018)	(0.029)	(0.198)
Month and year fixed effects			X
Judge and city fixed effects			X
Additional risk scores			X
Age and sex controls			X
Criminal history controls			X
Charge fixed effects			X
Observations	2654	2654	2654
R^2	0.016	0.100	0.278
Adjusted R^2	0.015	0.099	0.256
Residual std. error	0.490	0.468	0.425

Note: Even though PTS recommends detention for Black defendants at a higher risk-adjusted rate than white defendants (Table A9), we do not observe a risk-adjusted disparity in rates of release of Black defendants after the PTS investigation. We do not obtain statistically significant estimates of disparate impact for Hispanic defendants, or disparate treatment of either Black or Hispanic defendants.

Abbreviations: PTRA, Federal Pretrial Risk Assessment; PTS, pretrial services.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE A11 Racial disparate impact coefficients and standard errors derived from linear regression with linear, quadratic, and cubic terms for the PTRA criminal history score.

Outcome	Order	Hispanic–white disparity	Black–white disparity
Released at initial hearing	Linear	−0.132 (0.015)	−0.027 (0.015)
Released at initial hearing	Quadratic	−0.096 (0.015)	−0.007 (0.015)
Released at initial hearing	Cubic	−0.095 (0.015)	−0.004 (0.015)
Released after initial hearing	Linear	0.002 (0.023)	−0.014 (0.022)
Released after initial hearing	Quadratic	−0.005 (0.023)	−0.019 (0.022)
Released after initial hearing	Cubic	−0.007 (0.023)	−0.018 (0.022)
Released at any point	Linear	−0.059 (0.017)	−0.016 (0.016)
Released at any point	Quadratic	−0.058 (0.017)	−0.015 (0.016)
Released at any point	Cubic	−0.055 (0.017)	−0.011 (0.016)
AUSA moves for detention	Linear	0.12 (0.016)	0.042 (0.015)
AUSA moves for detention	Quadratic	0.088 (0.016)	0.024 (0.015)
AUSA moves for detention	Cubic	0.087 (0.016)	0.021 (0.015)
Consent before investigation	Linear	−0.011 (0.017)	−0.024 (0.016)
Consent before investigation	Quadratic	−0.012 (0.017)	−0.024 (0.017)
Consent before investigation	Cubic	−0.012 (0.017)	−0.024 (0.017)
Pretrial services rec. release	Linear	−0.008 (0.025)	−0.054 (0.024)
Pretrial services rec. release	Quadratic	−0.017 (0.025)	−0.062 (0.024)
Pretrial services rec. release	Cubic	−0.02 (0.025)	−0.06 (0.024)
Released after investigation	Linear	−0.009 (0.025)	−0.023 (0.024)
Released after investigation	Quadratic	−0.016 (0.025)	−0.029 (0.024)
Released after investigation	Cubic	−0.018 (0.025)	−0.027 (0.024)

Note: The observed disparities for release at any point are qualitatively similar across polynomial orders. For the outcomes of release at the initial hearing and the AUSA motion for detention, the Hispanic–white disparity shrinks by approximately 3pp in the higher order regressions, but remains statistically significant at the 95% level. The linear terms are duplicated in prior tables, but are included here for convenience.
Abbreviations: AUSA, assistant US attorney; PTRA, Federal Pretrial Risk Assessment.

TABLE A12 For the outcome of release at the initial hearing, comparison of disparate impact analyses across different risk scores

	Outcome: Released at initial hearing				
	Unadjusted	**Disparate impact**			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Hispanic–white disparity	−0.297***	−0.132***	−0.108***	−0.198***	−0.183***
	(0.017)	(0.015)	(0.015)	(0.016)	(0.015)
Black–white disparity	−0.286***	−0.027	−0.014	−0.074***	−0.042**
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
PTRA criminal history		−0.135***			
		(0.003)			
PTRA score			−0.090***		
			(0.002)		
PSA NCA				−0.051***	
				(0.004)	
PSA FTA				0.005	
				(0.006)	
PSA NVCA				−0.076***	
				(0.005)	
Estimated risk probabilities					−8.380***
					(0.206)
Constant	0.532***	0.916***	0.987***	0.725***	1.230***
	(0.011)	(0.013)	(0.013)	(0.011)	(0.020)
Observations	4809	4809	4809	4809	4809
R^2	0.088	0.333	0.355	0.303	0.313
Adjusted R^2	0.088	0.333	0.354	0.302	0.313
Residual std. error	0.458	0.391	0.385	0.400	0.397

Note: For Hispanic defendants, risk-adjusted disparities in rates of release at the initial hearing persist across risk scores, with a smaller effect for PTRA scores and a larger effect for PSA scores and estimated risk probabilities. For Black defendants, the pattern is similar, with smaller and statistically insignificant effects for PTRA scores and larger, statistically significant effects for PSA scores and and estimated risk probabilities.

Abbreviations: FTA, failures to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

*p < 0.05; **p < 0.01; ***p < 0.001.

TABLE A13 For the outcome of release after the initial hearing, comparison of disparate impact analyses across different risk scores

	Outcome: Released after initial hearing				
	Unadjusted	**Disparate impact**			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Hispanic–white disparity	−0.048*	0.002	−0.008	−0.022	−0.018
	(0.023)	(0.023)	(0.023)	(0.022)	(0.022)
Black–white disparity	−0.116***	−0.014	−0.043	−0.007	0.007
	(0.021)	(0.022)	(0.022)	(0.021)	(0.022)
PTRA criminal history		−0.077***			
		(0.005)			
PTRA score			−0.037***		
			(0.004)		
PSA NCA				−0.042***	
				(0.006)	
PSA FTA				−0.012	
				(0.008)	
PSA NVCA				−0.031***	
				(0.007)	
Estimated risk probabilities					−5.570***
					(0.290)
Constant	0.429***	0.728***	0.683***	0.650***	0.977***
	(0.017)	(0.027)	(0.031)	(0.020)	(0.034)
Observations	3042	3042	3042	3042	3042
R^2	0.010	0.072	0.041	0.108	0.104
Adjusted R^2	0.009	0.071	0.040	0.106	0.103
Residual std. error	0.480	0.464	0.472	0.456	0.456

Note: Across risk scores, we do not obtain statistically significant estimates of disparate impact for release after the initial hearing, though a directional disparity of 4pp exists for Black defendants after adjustment for PTRA scores (SE = 2pp, $p = 0.05$).

Abbreviations: FTA, failures to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE A14 For the outcome of release at any point in the pretrial process, comparison of disparate impact analyses across different risk scores

	Outcome: Released at any point				
	Unadjusted	**Disparate impact**			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Hispanic–white disparity	−0.209***	−0.059***	−0.058***	−0.105***	−0.092***
	(0.018)	(0.017)	(0.017)	(0.016)	(0.016)
Black–white disparity	−0.250***	−0.016	−0.032*	−0.031*	0.0003
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
PTRA criminal history		−0.122***			
		(0.003)			
PTRA score			−0.072***		
			(0.002)		
PSA NCA				−0.055***	
				(0.005)	
PSA FTA				−0.007	
				(0.007)	
PSA NVCA				−0.065***	
				(0.007)	
Estimated risk probabilities					−8.610***
					(0.215)
Constant	0.729***	1.080***	1.090***	0.930***	1.450***
	(0.010)	(0.011)	(0.012)	(0.009)	(0.018)
Observations	4809	4809	4809	4809	4809
R^2	0.054	0.245	0.216	0.271	0.280
Adjusted R^2	0.054	0.244	0.215	0.270	0.279
Residual std. error	0.479	0.428	0.436	0.420	0.418

Note: For Hispanic defendants, risk-adjusted disparities in rates of release at any point persist across risk scores, with a similar effect adjusting for PTRA scores and larger effects for PSA scores and estimated risk probabilities. For Black defendants, there are larger and statistically significant effects after adjusting for PTRA scores or PSA scores.

Abbreviations: FTA, failures to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

*p < 0.05; **p < 0.01; ***p < 0.001.

TABLE A15 For the outcome of whether or not the AUSA moves for detention, comparison of disparate impact analyses across different risk scores

	Outcome: AUSA moves for detention				
	Unadjusted	Disparate impact			
	(1)	(2)	(3)	(4)	(5)
Hispanic–white disparity	0.282***	0.120***	0.097***	0.185***	0.169***
	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)
Black–white disparity	0.295***	0.042**	0.030*	0.087***	0.055***
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
PTRA criminal history		0.132***			
		(0.003)			
PTRA score			0.088***		
			(0.002)		
PSA NCA				0.048***	
				(0.005)	
PSA FTA				−0.002	
				(0.006)	
PSA NVCA				0.076***	
				(0.006)	
Estimated risk probabilities					8.270***
					(0.209)
Constant	0.451***	0.074***	0.006	0.260***	−0.240***
	(0.011)	(0.013)	(0.013)	(0.011)	(0.020)
Observations	4809	4809	4809	4809	4809
R^2	0.086	0.316	0.334	0.289	0.301
Adjusted R^2	0.085	0.315	0.334	0.289	0.300
Residual std. error	0.463	0.401	0.395	0.408	0.405

Note: For both Hispanic and Black defendants, risk-adjusted disparities in the rates at which the AUSA moves for detention persist across risk scores, with a smaller effect adjusting for PTRA scores and larger effects for PSA scores and estimated risk probabilities.

Abbreviations: AUSA, assistant US attorney; FTA, failures to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE A16 For the outcome of consent at any point before the pretrial services investigation, comparison of disparate impact analyses across different risk scores

	Outcome: Consent before investigation				
	Unadjusted	**Disparate impact**			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Hispanic–white disparity	−0.018	−0.011	0.0003	−0.023	−0.015
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Black–white disparity	−0.037*	−0.024	−0.004	−0.036*	−0.026
	(0.015)	(0.016)	(0.016)	(0.016)	(0.016)
PTRA criminal history		−0.010*			
		(0.004)			
PTRA score			−0.017***		
			(0.003)		
PSA NCA				0.010*	
				(0.004)	
PSA FTA				−0.010	
				(0.006)	
PSA NVCA				−0.011	
				(0.006)	
Estimated risk probabilities					−0.510*
					(0.227)
Constant	0.157***	0.195***	0.270***	0.158***	0.207***
	(0.012)	(0.020)	(0.024)	(0.015)	(0.026)
Observations	3042	3042	3042	3042	3042
R^2	0.002	0.004	0.014	0.004	0.004
Adjusted R^2	0.001	0.003	0.013	0.002	0.003
Residual std. error	0.343	0.342	0.341	0.343	0.343

Note: After adjusting for PSA scores, Black defendants are 4pp *less* likely than similarly risky white defendants to consent to detention before the PTS investigation (SE = 2pp, $p = 0.02$), though we do not obtain statistically significant estimates of disparate impact for other risk scores. Across risk scores, we do not obtain statistically significant estimates for disparate impact for Hispanic defendants.

Abbreviations: FTA, failures to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE A17 For the outcome of whether or not pretrial services recommends release, comparison of disparate impact analyses across different risk scores

	Outcome: Pretrial services recommends release				
	Unadjusted	**Disparate impact**			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Hispanic–white disparity	−0.066**	−0.008	−0.012	−0.035	−0.029
	(0.025)	(0.025)	(0.025)	(0.024)	(0.024)
Black–white disparity	−0.178***	−0.054*	−0.080***	−0.049*	−0.027
	(0.023)	(0.024)	(0.024)	(0.023)	(0.023)
PTRA criminal history		−0.091***			
		(0.006)			
PTRA score			−0.050***		
			(0.004)		
PSA NCA				−0.050***	
				(0.006)	
PSA FTA				−0.012	
				(0.008)	
PSA NVCA				−0.031***	
				(0.008)	
Estimated risk probabilities					−6.550***
					(0.299)
Constant	0.501***	0.859***	0.849***	0.748***	1.150***
	(0.018)	(0.028)	(0.033)	(0.021)	(0.034)
Observations	2654	2654	2654	2654	2654
R^2	0.024	0.108	0.077	0.147	0.150
Adjusted R^2	0.023	0.107	0.076	0.145	0.149
Residual std. error	0.486	0.464	0.472	0.454	0.453

Note: After adjusting for PTRA criminal history scores, PTRA scores, or PSA scores, PTS is less likely to recommend release for Black defendants, though we do not obtain statistically significant estimates of disparate impact in the rates at which PTS recommends detention for Hispanic defendants.

Abbreviations: FTA, failures to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE A18 For the outcome of release at any point after the pretrial services investigation, comparison of disparate impact analyses across different risk scores

	Outcome: Released after investigation				
	Unadjusted	**Disparate impact**			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Hispanic–white disparity	−0.067**	−0.009	−0.012	−0.041	−0.031
	(0.025)	(0.025)	(0.025)	(0.024)	(0.024)
Black–white disparity	−0.148***	−0.023	−0.048*	−0.019	0.001
	(0.023)	(0.024)	(0.024)	(0.023)	(0.023)
PTRA criminal history		−0.092***			
		(0.006)			
PTRA score			−0.051***		
			(0.004)		
PSA NCA				−0.042***	
				(0.006)	
PSA FTA				−0.016	
				(0.009)	
PSA NVCA				−0.042***	
				(0.008)	
Estimated risk probabilities					−6.490***
					(0.308)
Constant	0.500***	0.860***	0.854***	0.746***	1.140***
	(0.018)	(0.029)	(0.034)	(0.021)	(0.035)
Observations	2654	2654	2654	2654	2654
R^2	0.016	0.100	0.071	0.135	0.139
Adjusted R^2	0.015	0.099	0.069	0.133	0.138
Residual std. error	0.490	0.468	0.476	0.459	0.458

Note: Adjusting for PTRA scores, Black defendants are 5pp less likely to be released after the PTS investigation than similarly risky white defendants (SE = 2pp, $p = 0.04$), but the effect is not statistically significant for other risk scores. Across risk scores, we do not obtain statistically significant estimates of disparate impact for Hispanic defendants.

Abbreviations: FTA, failures to appear; NCA, new criminal activity; NVCA, new violent criminal activity; PSA, Public Safety Assessment; PTRA, Federal Pretrial Risk Assessment.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.