Striking a Balance: Human Discretion and Algorithmic Insights in Parole Supervision Decision-Making

Anuar Assamidanov* Nicholas Powell†

October 2023

Abstract

In this paper, we examine the interplay between predictive algorithms and human discretion in determining parole supervision levels. Adopting a methodological approach centered on the random assignment of parole officers at specific risk score thresholds—particularly at junctures where parolees transition between various supervision levels—we investigate the impact of officers' decisions to deviate from algorithmic recommendations on recidivism rates. Our findings reveal that professional adjustments to higher supervision levels consistently lead to reduced recidivism rates, while adjustments to lower supervision levels don't display a significant effect. This underscores the pivotal role of strategic resource allocation in parole supervision, indicating that harsh overrides can be resource-optimal in effectively lowering recidivism. Conversely, lenient overrides maintain stable recidivism rates without necessitating intensified supervision. Additionally, the study contributes to the ongoing discourse on the role of human intervention in algorithmic recommendations within the criminal justice system.

*Department of Economics, Claremont Graduate University, anuar.assamidanov@cgu.edu

†Georgia Department of Community Supervision, Director of Strategic Planning & Analysis.
We would like to express our deepest gratitude to Gregory DeAngelo, Fernando Lozano, Mark Hoekstra, Rebecca Thornton, Bryan McCannon, Sonja Starr and the participants of various seminars for their invaluable feedback and constructive critiques. Any shortcomings in this paper are entirely our own. We also extend thanks to the participants of the 16th All-California Labor Economics Conference, the 17th Annual Conference on Empirical Legal Studies, and the Workshop Convening of Prosecutor-Researcher Collaborations for their insightful input. A special acknowledgment goes to the Georgia Department of Community Supervision for not only providing the essential data for our study but also for their meaningful discussions and critical insights that significantly enriched our work.

The last decade has seen an upsurge in the use of predictive algorithms in various critical domains, including job screening, medical diagnoses, and pretrial release decisions (Obermeyer et al., 2019). These algorithms, driven by the enormous potential of artificial intelligence and big data, aim to reduce human error and increase efficiency in decision-making processes (Dietvorst et al., 2015). Despite the growing reliance on algorithmic systems, the final decision-making authority often remains in human hands, believing that human oversight can provide valuable insights and rectify algorithmic inaccuracies (Dressel and Farid, 2018). In criminal justice settings, professional discretion has long been recognized as one of the foundational elements of effective risk tool implementation (Andrews et al., 1990).

Therefore, this study explores the complex interplay between human discretion and algorithm-based decision-making in parole supervision decisions. Our investigation situates itself at the crossroads of two major bodies of literature. On the one hand, there is significant work examining the role of algorithmic prediction in the criminal legal system, which delves into the ethical and practical implications of relying on data-driven models for policing, sentencing, and bail decisions (Kleinberg et al., 2018; Stevenson and Doleac, 2021; Stevenson, 2017; Angelova et al., 2023). On the other hand, there is an extensive research history scrutinizing the effectiveness of community supervision, focusing on the impact of different supervision models and intensities on outcomes such as recidivism and social reintegration (Piehl and LoBuglio, 2005; Hawken and Kleiman, 2009; Rose, 2021; Sakoda, 2023). While each area offers valuable insights, there remains a gap in understanding how they intersect, particularly concerning the role and impact of human discretion in algorithm-advised decisions. This study aims to bridge this gap by examining how parole officers' discretion influences the outcomes of decisions guided by predictive algorithms.

Parole officers, equipped with algorithm-generated risk scores that consider variables such as criminal history, age, and social support, have the latitude to override these recommendations based on additional information or perceived shortcomings of the algorithm (Monahan and Skeem, 2016). While this human intervention has the potential to enhance or compromise the effectiveness of the algorithm-based system, the empirical evidence supporting either perspective remains sparse and inconclusive.

The introduction of predictive algorithms in parole supervision decisions has sparked an important debate in the realms of behavioral science and criminology. While these algorithms can increase efficiency and minimize human error, there is an ongoing discussion about the role and impact of human discretion in these decisions. This study contributes to this debate by exploring how parole officers' discretion influences the outcomes of algorithm-based decisions (Harcourt, 2007).

One central aspect of this exploration is the comparison of observed supervision levels assigned by parole officers and hypothetical supervision levels suggested by the algorithm. Such a comparison can illuminate whether parole officers' decisions provide valuable insights that enhance the algorithmic recommendations or introduce biases and inaccuracies (Dietvorst et al., 2015). However, this approach faces a significant selection challenge, given that recidivism rates can only be observed for parolees who were assigned specific supervision levels by parole officers. This poses a difficulty in measuring the counterfactual outcomes of the alternative supervision level that was not chosen (Berk, 2017).

To address the selection challenge, we exploited the random assignment of parole officers to parolees, assessing the causal impact of overrides on recidivism rates. Specifically, we conditioned on risk scores near thresholds delineating different supervision levels. This approach adeptly manages missing data, enabling estimation of potential outcomes across both observed and unobserved supervision levels (Berk et al., 2018b). Our study draws from data on individuals released on discretionary parole to the Georgia Department of Community Supervision (DCS) from January 1, 2013, to December 31, 2015. This dataset, comprising over 25,000 de-identified records, offers insights into demographics, prison and parole details, prior community supervision histories, supervision conditions, and recidivism rates. The comprehensive analysis of this rich dataset serves as the core of our investigation.

The findings indicate that overrides, where parole officers exercise their discretion to deviate from the algorithm's recommended supervision level, significantly affect recidivism rates. Specifically, we found that when parole officers decide to assign a higher supervision level than suggested by the algorithm, recidivism rates significantly decrease. This suggests that in some instances, human intuition and expertise can identify factors not considered by the algorithm, leading to better-informed decisions and improved outcomes (Dressel and Farid, 2018). However, when officers chose to lower the supervision level contrary to the algorithm's recommendation, we did not find any significant effect on recidivism rates. This reveals that reduced supervision in these instances doesn't necessarily translate to a higher likelihood of re-offending. In other words, similar results with fewer agency resources.

Building on these findings, the remainder of this paper will discuss their implications for the criminal justice system. We propose that a more balanced approach, combining the strengths of human decision-makers and predictive algorithms, can lead to better outcomes. Achieving this balance will require not only improved training and education programs for parole officers but also the refinement of predictive algorithms to consider more contextual factors and individual differences. Such adjustments can potentially contribute to a more equitable and accurate decision-making process (Monahan and Skeem, 2016). Through this work, we aim to shed light on the interplay between human discretion and algorithmic predictions, ultimately informing more effective oversight policies and promoting fairer, more accurate outcomes in the criminal justice system.

The remainder of this paper is structured as follows: First, in Section 1, we contextualize the role of algorithmic risk scores and overrides in parole supervision. This sets the stage for Section 2, which presents the characteristics of parole cases segmented by the types of supervision decisions made. We then outline our methodology in Section 3, detailing the empirical strategy that combines threshold-based methods with the random assignment of parole officers to parolees. Following this, we present our findings in Section 4 and delve into their broader implications in Section 5. Through this comprehensive approach, we aim to provide a nuanced understanding of the complex interplay between human discretion and algorithmic decision-making in parole supervision.

1 Background and Context

The criminal justice system serves the intricate balance of protecting societal safety and respecting the rights and rehabilitation of individuals who have served their sentences for criminal offenses. Parole officers function as pivotal actors within this multifaceted system. They are entrusted with the supervision of individuals who have been released from incarceration and the responsibility of ensuring their adherence to the conditions of their parole (Paparozzi and Gendreau, 2005).

Parole supervision in Georgia is articulated around three primary components. The first is the supervision intensity level, which stipulates the frequency and mode (e.g., in-person, over the phone) by which a person must report to their parole officer. Next, there are specific conditions that individuals must comply with during the duration of their parole. Such conditions can encompass requirements like curfews, mandatory drug and alcohol treatments, contact restrictions, or even placement in a community corrections center or halfway house. The last component is the assignment of a parole officer. This officer plays a vital role in ensuring compliance with the preceding components, linking individuals to resources, and wielding considerable discretion in shaping the parolee-officer relationship.

Parole officers bear the critical task of ascertaining the intensity of supervision services for each person, ranging from intensive monitoring and resource linkage to more lenient oversight. Ideally, this supervision level mirrors people's risk of re-offending or recidivism - an evaluation typically facilitated through risk assessment tools (Taxman, 2002).

One prevalent tool is the algorithm-generated risk score. This score is computed based on various factors, including but not limited to the individual's criminal history, age, employment status, and social support system. The score then predicts a person's likelihood of re-offending (Berk et al., 2018a). Leveraging this risk score, the algorithm proposes a supervision level designed to minimize the recidivism risk while optimizing the allocation of parole supervision resources (Monahan and Skeem, 2016).

However, parole officers maintain the discretion to override these algorithmic recommendations. Overrides might occur when the officer possesses additional information not captured by the algorithm or believes the algorithm lacks a nuanced understanding of a person'sunique circumstances (Dietvorst et al., 2015). This intersection of algorithmic recommendations and human discretion forms the crux of our current investigation.

The parole supervision process often faces challenges due to the non-random assignment of parolees to parole officers. Such a non-randomized approach can lead to selection bias, as the supervision styles and override tendencies of individual parole officers might significantly influence the outcomes for the parolees under their charge (Berk, 2017). Recognizing this limitation in existing studies, our research introduces a novel approach by implementing a randomized assignment of parole officers to parolees. This methodological advancement is crucial in our study, as it enables a more rigorous and unbiased analysis of the effects of human discretion on parole decisions. By ensuring a randomized allocation, we mitigate the potential biases that could arise from the idiosyncrasies of parole officers, thereby providing a stronger foundation for assessing the true impact of parole officer decisions on recidivism rates.

In this study, as illustrated in Figure 1, the algorithm-generated risk score is numerical, ranging from 1 to 10. Each score range corresponds to a recommended supervision level. Specifically, a risk score between 1 and 5 suggests standard supervision for those with a lower recidivism risk. Scores between 6 and 8 call for high supervision, typically for those with a substantial criminal history or other risk factors. Finally, scores between 9 and 10 mandate specialized supervision for those with a severe criminal history or elevated risk factors.

Figure 1. Algorithmic Risk Score and Corresponding Supervision Level Recommendations

Supervision Level	Standard					High			Specialized	
Algorithmic Risk Score	1	2	3	4	5	6	7	8	9	10

Notes: Figure based on the framework developed by the Community Supervision Department in Georgia for the allocation of parolees to respective supervision levels based on their algorithmic risk scores.

Diving deeper into the dynamics of parole decisions, our study highlights the crucial concept of overrides. An override occurs when a parole officer, drawing from their professional expertise or ancillary information beyond the algorithm's scope, opts for a supervision level divergent from the algorithmic risk score's recommendation. Specifically, there are two principal override types: the "Harsh Override," where a parole officer, upon evaluation of supplementary factors, recommends a more stringent supervision level than the algorithm, and the "Lenient Override," where more lenient supervision is deemed appropriate.

In the realm of parole decisions, while algorithms provide standardized suggestions rooted in comprehensive data, parole officers incorporate the multifaceted realities and idiosyncrasies inherent in each case. This contrast between consistent algorithmic advice and the detailed human decision-making process is central to our research. The scope and consequences of these discretionary overrides are central to our research exploration.

2 Data

The data for this study was sourced from the State of Georgia, focusing on individuals released from Georgia prisons on discretionary parole to the Georgia Department of Community Supervision (DCS) between January 1, 2013, and December 31, 2015. These individuals were under post-incarceration supervision, and the data included various aspects of their supervision and prior history.

The DCS provided comprehensive data on the supervised individuals, encompass-ing demographic information, prison and parole case details, prior community supervision history, and the conditions of supervision set by the Board of Pardons and Paroles. Further, the data included records of supervision activities such as violations, drug tests, program attendance, employment, residential moves, and accumulation of delinquency reports for violating parole conditions.

In addition to the data provided by the DCS, the Georgia Bureau of Investigation supplied data from the Georgia Crime Information Center (GCIC), a statewide repository of criminal history records. The GCIC data offered a detailed account of the individual's prior criminal history in Georgia, including past arrests and convictions prior to prison entry. The GCIC data also provided our key measure of recidivism, defined as a new felony or misdemeanor arrest within three years of the parole supervision start date.

However, approximately 8% of the original population was excluded due to various reasons such as lack of a unique identifier to link DCS to GCIC data, invalid Georgia zip code, transfer to another state for supervision, invalid birth date, or death. Youths under the age of 18 at the time of prison release were also excluded.

The final dataset, after these exclusions, consisted of over 25,000 de-identified records. These records were devoid of personal, address, and agency identifiers to protect the privacy of the individuals. To prevent potential deductive disclosure, the data included only two racial categories: Black and White.

The data was further enhanced by pairing it with information from the U.S. Census Bureau's Public Use Microdata Area (PUMA). Each individual's residential address at the time of prison release was mapped to a PUMA, and neighboring PUMAs were grouped into 25 unique spatial units.

This rich, multi-faceted dataset provides a valuable resource for examining the role of human discretion in parole supervision decisions, and the impact of such discretion on recidivism rates.

Table 2 provides an analysis of our evaluation sample, now selectively comprised of cases at the margins of different supervision levels. This targeted approach is designed to enhance the comparability of observable characteristics within our sample. In Panel A, we observe that males continue to dominate the sample, accounting for 87% of all cases. The proportion of males increases to 93% in the subset where parole officers elected a harsh override, suggesting a tendency towards stricter supervision for male parolees. Meanwhile, white individuals are predominantly present in the group that follows the algorithm's recommendation for a high supervision level at 43%, while they are less common at 35% in the group receiving a harsh override, indicating a potential discrepancy in treatment by race.

Panel B provides insight into the prior criminal history of the parolees. Notably, individuals who received harsh overrides have a lower average number of felony convictions (1.35) and drug-related arrests (1.26) compared to those given lenient overrides, with averages of 1.47 and 2.02, respectively. The same trend is observed in property arrest episodes, where those with harsh overrides have fewer arrests (1.88) than their lenient counterparts (2.45). These patterns could indicate that parole officers are taking into account additional contextual information that may not be fully captured by the algorithm. This suggests a nuanced decision-making process where officers may consider the severity or nature of past offenses and other mitigating factors in their supervision-level recommendations.

Panel C discusses the factors that the algorithm considers. Individuals who received harsh overrides tend to be older at the time of release and have served longer prison sentences, particularly those without property offenses, suggesting a focus on more stringent supervision for those with potentially higher reintegration challenges. Conversely, lenient overrides are often granted to younger individuals with a higher number of felony and misdemeanor arrests, including more prior revocations of parole and probation, indicating a tendency towards providing opportunities for less restrictive supervision despite a history of non-compliance. These patterns underscore the parole officers' nuanced decision-making, taking into account not just the risk of recidivism, but also the rehabilitative needs and the severity of past offenses.

Finally, Panel D highlights the recidivism outcomes. Parolees subjected to harsh overrides, which intensify their supervision level against the algorithm's recommendation, exhibit a lower recidivism rate within 3 years at 55%, suggesting that increased supervision may mitigate re-offending. However, those who benefit from lenient overrides show a slightly higher recidivism rate of 63% within the same period. This could indicate that, in some cases, reduced supervision does not necessarily increase the likelihood of re-offending, which may reflect the successful integration of parolees back into the community or other unmeasured factors.

Table 1. Summary Statistics of Parole Case Characteristics by Supervision Decision within Threshold

	All Cases	Follow Algorithm	Harsh Override	Lenient Override
Demographics				
Male	0.87	0.86	0.93	0.86
White	0.42	0.43	0.35	0.42
Prior Criminal History				
Prior_Conviction_Episodes_Felony	1.36	1.34	1.35	1.47
Prior_Conviction_Episodes_Misd	1.71	1.71	1.59	1.79
Prior_Conviction_Episodes_Viol	0.32	0.29	0.49	0.32
Prior_Arrest_Episodes_Property	2.18	2.18	1.88	2.45
Prior_Arrest_Episodes_Drug	1.84	1.89	1.26	2.02
Prior_Arrest_Episodes_PPViolationCharges	2.35	2.35	1.97	2.6
Prior_Arrest_Episodes_DVCharges	0.16	0.16	0.18	0.17
Prior_Arrest_Episodes_GunCharges	0.27	0.27	0.28	0.29
Prior_Conviction_Episodes_Prop	1.1	1.09	0.94	1.26
Prior_Conviction_Episodes_Drug	0.79	0.82	0.53	0.85
Prior_Conviction_Episodes_PPViolationCharges	0.32	0.32	0.31	0.33
Prior_Conviction_Episodes_DomesticViolenceCharges	0.08	0.08	0.09	0.07
Prior_Conviction_Episodes_GunCharges	0.14	0.13	0.13	0.17
Algorithmic Inputs				
Age_at_Release	32.03	31.85	33.02	32.15
Prison_Years	1.78	1.66	2.49	1.83
Prison_Offense_Property	0.34	0.35	0.23	0.36
Prior_Arrest_Episodes_Felony	5.65	5.61	5.08	6.24
Prior_Arrest_Episodes_Misd	3.26	3.27	2.95	3.39
Prior_Arrest_Episodes_Violent	0.97	0.92	1.32	0.98
Prior_Revocations_Parole	0.09	0.08	0.08	0.14
Prior_Revocations_Probation	0.15	0.15	0.1	0.18
Recidivism Outcomes				
Recidivism_Within_3years	0.61	0.62	0.55	0.63
Recidivism_Arrest_Year1	0.32	0.32	0.27	0.32
Recidivism_Arrest_Year2	0.19	0.19	0.19	0.2
Recidivism_Arrest_Year3	0.11	0.11	0.08	0.12
Cases	8352	6157	1254	941

Notes: This table presents the summary statistics of parole case characteristics for a subset of the sample at the threshold of changing supervision levels, enhancing comparability across groups. This refined approach allows for a more precise assessment of the impact of parole officer discretion on recidivism outcomes. The sample includes 8,352 cases from Georgia prisons, with individuals released on discretionary parole to the Georgia Department of Community Supervision (DCS) between January 1, 2013, and December 31, 2015. Cases are segmented based on whether parole officers followed the algorithm's recommendations or exercised discretion through harsh or lenient overrides. All reported values are means for the variables indicated.

In continuing our analysis, Table 2 presents a systematic breakdown of recidivism rates associated with parole supervision decisions, both as recommended by an algorithm and when overridden by parole officers. Recidivism rates are stratified based on the time post-release: within three years and separately for each year within that timeframe.

In the High recommendation category, following the algorithm yields a recidivism rate of 64% within three years. However, we notice discernible differences when parole officers exercise discretion and override the recommendation. Lenient overrides lead to a slightly decreased rate of 61.5% within three years, whereas harsh overrides see a more substantial reduction to 57.4%. This trend of decreased recidivism with harsh overrides compared to algorithm-following is similarly observed in the Standard recommendation category. In contrast, the Specialized category exhibits the highest recidivism rate when strictly adhering to the algorithm at 69.3%. Lenient overrides in this category bring the rate down to 67.0%.

Table 2. Recidivism Rates at Supervision Borderline by Algorithmic Recommendation and Override Type

			Recidivism Arrest			
Recommend	Override	Cases	Within 3 Years	Year 1	Year 2	Year 3
High	Follow Algorithm	2979	0.633	0.331	0.188	0.115
	Harsh Override	570	0.563	0.293	0.182	0.088
	Lenient Override	943	0.624	0.314	0.187	0.123
Specialized	Follow Algorithm	1407	0.681	0.369	0.201	0.111
	Lenient Override	311	0.666	0.347	0.222	0.096
Standard	Follow Algorithm	1771	0.545	0.276	0.178	0.091
	Harsh Override	371	0.518	0.240	0.202	0.075

Notes: This table presents recidivism rates for parolees at the threshold of changing supervision levels, segmented by supervision level (High, Specialized, Standard) and decision type (follow algorithm, harsh override, lenient override). It shows rates for overall recidivism within 3 years, and separately for each of the first three years post-release. The 'Follow Algorithm' category represents cases where parole officers adhered to the algorithm's recommendation. 'Harsh Override' indicates a stricter supervision level than recommended, while 'Lenient Override' refers to a more lenient level than suggested by the algorithm. This focused analysis at the threshold provides a clearer understanding of the impact of supervision decisions on recidivism in borderline cases.

Year-wise, the first year consistently exhibits the highest recidivism rates across all categories and decision types, with subsequent years showing a gradual decline.

For instance, within the High recommendation category, the first-year recidivism rate stands at 34% when following the algorithm, decreasing slightly to 31.1% with lenient overrides and further to 28.2% with harsh overrides. This trend underscores the influence of supervision decisions, whether algorithmic or human-driven, on the early post-release period and its long-term implications.

The findings from this table underline the varying effectiveness of algorithmic recommendations and discretionary overrides in mitigating recidivism, segmented by risk categories. The nuanced impact of these approaches across the different risk categories warrants further investigation. In the next section, we will delve into a deeper analysis of these findings, seeking to understand the dynamics behind these patterns

3 Methodology

The objective of this research is to meticulously examine the effects of parole officers' discretion to override predetermined algorithmic risk scores on recidivism rates among parolees. The crux of our investigation rests on instances when the designated supervision level - whether standard, high, or specialized - is subject to change due to an override decision made by a parole officer. We are particularly interested in understanding whether these discretionary decisions, which veer away from the algorithmic recommendations, yield a significant influence on recidivism rates.

In our analytical pursuit, we utilize a methodological approach that specifically relies on arbitrary cutoff points, commonly referred to in this context as threshold-based analysis. This methodology is particularly apt for our research, given its inherent emphasis on these predetermined cutoffs. Within our study's framework, these thresholds translate to risk scores that dictate a change in the level of supervision. For instance, a certain risk score may prompt a transition for a parolee from standard supervision to high supervision, or from high to specialized supervision.

The underlying assumption of this approach, which also forms the basis of its strength, posits that proximate to these thresholds, the allocation of individuals to distinct supervision levels approximates a quasi-random process. If we observe any significant deviation in recidivism rates at the threshold, we can confidently attribute this to the change in supervision level, which we refer to as the 'treatment' effect in this study.

Within our analytical framework, we commence by delineating threshold variables, which, in this context, pertains to the risk scores attributed to individuals. Specifically, we focus on individuals characterized by risk scores of 5, 6, 8, and 9, as depicted in Figure 1. These demarcated cutoff values are pivotal since they represent the specific risk score junctures at which there's a consequential shift in the supervision level allocated to the parolee.

In our analytical approach, we enhance conventional methods by leveraging the random assignment of parole officers to parolees. This unique allocation helps control for officer-specific attributes that could confound our results. With this foundation, we incorporate the decisions to override into our threshold analysis, specifically targeting the influence of parole officers' discretionary decisions on recidivism rates. This dual-faceted method allows us not only to assess the initial algorithmic recommendations but also to understand the implications of officers' decisions when they diverge from these algorithmic suggestions.

To ensure the validity of our causal interpretations, we conducted two separate balance checks for observable covariates. The first balance check compared individuals who received a harsh override to their supervision level against those who followed the algorithmic recommendation. The second balance check compared individuals who received a lenient override against the same baseline group. In both checks, we controlled for a comprehensive set of factors, including demographic characteristics, prior criminal history, and specific conditions of supervision. These balance checks are essential to determine whether parole officers' discretionary overrides are influenced by variables not accounted for by the algorithm, providing a robust basis for our subsequent analysis of recidivism outcomes. We modeled these comparisons using the following regression equations:

For harsh overrides:

$$HarshOverride_i = \beta_0 + \beta_1 X_{i,\text{demog}} + \beta_2 X_{i,\text{crime}} + \beta_3 X_{i,\text{supervision}} + \epsilon_i \tag{1}$$

For lenient overrides:

$$LenientOverride_i = \gamma_0 + \gamma_1 X_{i,\text{demog}} + \gamma_2 X_{i,\text{crime}} + \gamma_3 X_{i,\text{supervision}} + \zeta_i \tag{2}$$

Here, $HarshOverride_i$ and $LenientOverride_i$ indicate whether an individual i received a harsh or lenient override, respectively. $X_{i,\text{demog}}$ includes demographic variables such as gender, race, and education level, $X_{i,\text{crime}}$ represents prior criminal history including arrests and convictions, and $X_{i,\text{supervision}}$ captures the conditions of supervision. The coefficients β_1, β_2, β_3, and γ_1, γ_2, γ_3 measure the relationship between these covariates and the likelihood of receiving either type of override. The error terms ϵ_i and ζ_i capture unobserved factors affecting the override decision. A lack of significance would imply that overrides are not systematically biased with respect to observed covariates, reinforcing the validity of subsequent analyses of recidivism outcomes. If any coefficients are significant, it would indicate a potential imbalance, and the direction of the coefficient would suggest the direction of the bias. This detailed approach allows us to understand and control for any systematic differences between the overridden and non-overridden groups, thereby enhancing the credibility of our findings regarding the impact of overrides on recidivism.

Our study's methodology centers around regression analysis, meticulously crafted to discern the causal relationship between parole officers' override decisions and recidivism rates. At the core of our analysis is the primary regression equation:

$$Recidivism_i = \gamma_0 + \gamma_1 O_i + \gamma_2 O_i \times Harsh_i + \gamma_3 O_i \times Lenient_i + \gamma_4 Threshold_i + \gamma_5 X_i + \zeta_i \tag{3}$$

Within this framework, $Harsh_i$ and $Lenient_i$ serve as binary indicators, clarifying the nature of the override for each individual. The coefficients γ_2 and γ_3 emerge as crucial parameters, revealing the differential impacts of stringent versus accommodating override decisions on the likelihood of reoffending.

It is pivotal to break down the relative effects of the two types of overrides. Harsh overrides, by their nature, signify a belief that the algorithm's recommendation is too lenient, thereby necessitating stricter supervision. On the contrary, lenient overrides indicate the opposite; a parole officer believes the recommendation is too strict.

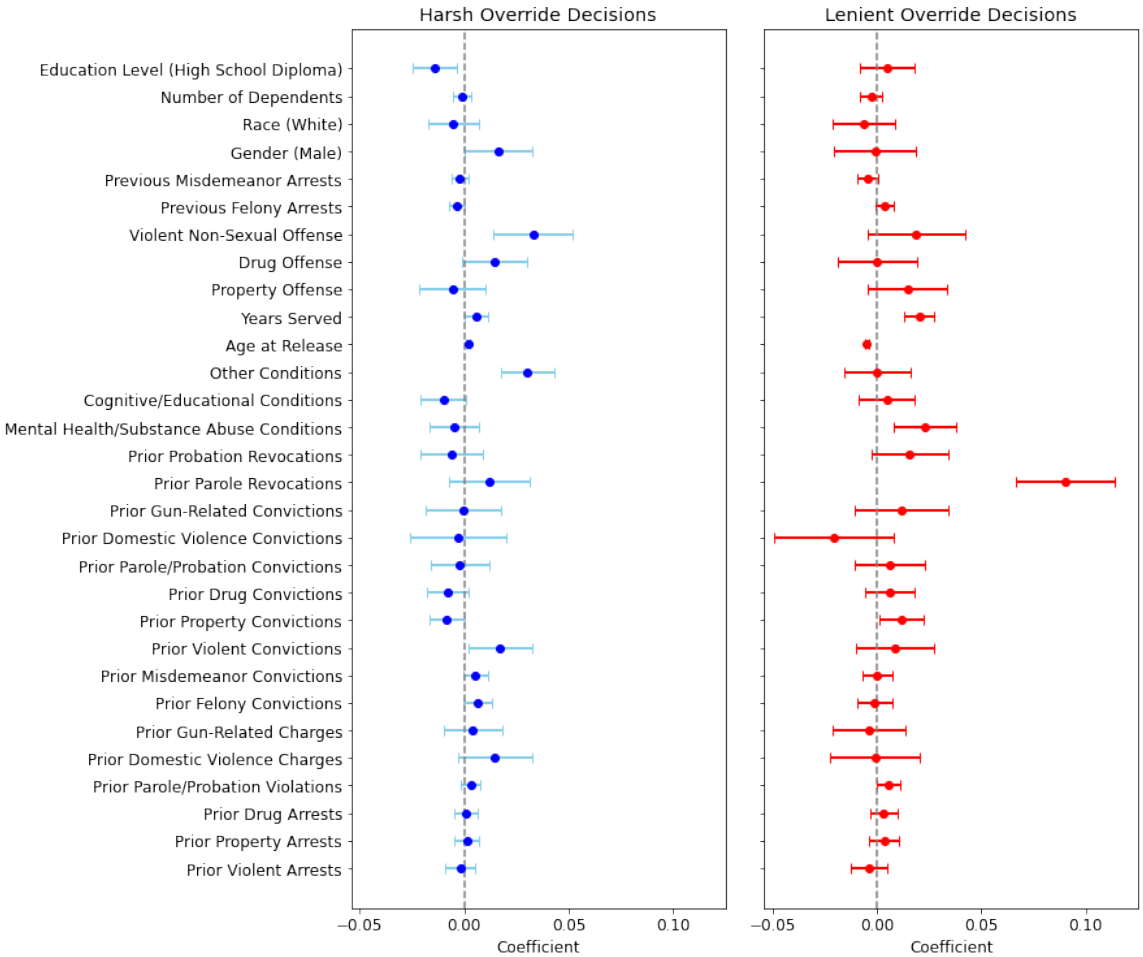
In the realm of recidivism, a harsh override could lead to an increased likelihood of reoffending due to the stringent conditions imposed, potentially creating an environment that doesn't favor reform. Conversely, lenient overrides, by relaxing certain conditions, might create a more favorable environment for the parolee, thus reducing the likelihood of recidivism. The actual outcomes, however, might vary, and this makes the coefficients γ_2 and γ_3 indispensable in determining the real-world effects of these decisions.

Concluding, our multifaceted methodological approach is primed to offer a comprehensive and nuanced assessment of the ramifications of parole officers' override decisions on recidivism rates. Through these meticulously designed regression models, we anticipate a thorough understanding, capturing a broad spectrum of variables and shedding light on the complexities inherent in parole decision-making.

4 Results

The empirical results are derived from the comprehensive analytical framework outlined in the previous section. We begin by evaluating the first equation that satisfies the critical assumption for the borderline method. The regression analysis used in the balance test, as depicted in Figure 2, reveals subtle distinctions in the characteristics of parolees who are subject to harsh overrides compared to those who are not. While most variables, including 'Prior Arrest Episodes Violent', 'Property', and 'Drug', yield coefficients straddling zero within their confidence intervals, they suggest no marked distinction in these aspects between the two groups. This indicates a baseline similarity in these factors, implying that the decision for harsh overrides is not influenced by these characteristics across the board.

Figure 2. Balance Test of Parolee Characteristics by Supervision Override Decisions



Notes: The balance test depicted here analyzes the relationship between various parolee characteristics and the likelihood of receiving either lenient or harsh supervision overrides as compared to standard algorithmic recommendations. The dataset encompasses a total of 8,352 parole cases, with 1,254 cases subject to harsh overrides and 941 to lenient overrides. The plotted coefficients, derived from regression models conditional on supervision level and geographic Public Use Microdata Area (PUMA), illustrate the magnitude and direction of the association between each characteristic and the supervision decision. Error bars represent 95% confidence intervals; coefficients for which intervals cross the zero line are not statistically significant, indicating no evidence of a difference in those characteristics between parolees given overrides and those who received standard algorithmic supervision.

However, certain variables like 'Prior Property Convictions' and 'Education Level High School Diploma' show negative coefficients with confidence intervals entirely below zero, albeit with a modest magnitude. This suggests that such factors might inversely correlate with the likelihood of receiving a harsh override, potentially indicating that parole officers view these attributes as less predictive of recidivism, warranting less intensive supervision. Conversely, attributes associated with higher risk and more serious offenses—such as 'Violent Non-Sexual Offense', 'Drug Offense', 'Prior Violent Convictions', along with the conditions of 'No Victim Contact', 'Electronic Monitoring', or 'Restitution', and the demographic indicator of being male—exhibit positive coefficients outside the zero bounds. This pattern hints at a more prevalent application of strict supervision measures among individuals with these characteristics, suggesting that parole officers might employ a more conservative approach when these risk indicators are present. Such a strategy may be indicative of parole officers exercising their judgment in a manner that aligns with a heightened assessment of recidivism risk, thereby tailoring their oversight to mitigate these concerns.

Furthermore, the overall landscape of lenient overrides, as depicted by the regression analysis, is complex and informed by a variety of factors. While the majority of variables show no discernible difference in treatment—evident in coefficients straddling zero, such as for 'Prior Arrest Episodes for Violent, Property, and Drug-related Offenses'—the factor 'Prior Revocations Parole' stands out with a notably high coefficient (0.09). This positive coefficient, firmly above zero, suggests that parolees with a history of parole revocations are more likely to be granted leniency, potentially as part of a deliberate strategy emphasizing rehabilitation over stricter supervision.

In conjunction with this, positive coefficients for 'Mental Health/Substance Abuse Conditions' (0.023) and 'Violent Non-Sexual Offence' (0.019) indicate that parole officers might favor lenient overrides for individuals with mental health or substance abuse issues and those with non-sexual violent offenses. This approach could signify a tailored response to the complex rehabilitation needs of these groups. In contrast, older parolees, indicated by the negative coefficient for 'Age at Release' (-0.0047), appear less likely to receive such leniency, pointing towards a perception that existing supervision protocols suffice or are more effective for this demographic.

Turning to equation 3, which estimates the causal effect of override on recidivism, Table 3 below provides a comprehensive overview. In this regression analysis, we use the new felony/misdemeanor within 3 years of supervision start as our outcome variable. We gradually include different sets of control variables in our model, ranging from demographics to supervision activities, enabling us to isolate the effect of the override decision on recidivism.

Table 3. Effects of Override on Recidivism Rates (with Harsh and Lenient Overrides)

	(1)	(2)	(3)	(4)	(5)
Outcome: New Felony/Mis within 3 Years of Supervision Start					
Harsh Override	-0.0738***	-0.0800***	-0.0604***	-0.0657***	-0.0519***
	(0.0134)	(0.0133)	(0.0133)	(0.0136)	(0.0137)
Lenient Override	0.0150	0.0107	-0.0048	-0.0102	-0.0111
	(0.0119)	(0.0118)	(0.0116)	(0.0116)	(0.0113)
Demographics		Y	Y	Y	Y
Criminal History			Y	Y	Y
Conditions of Supervision				Y	Y
Supervision Activities					Y
N	8352	8352	8352	8352	8125

Notes: This table presents the estimated effects of harsh and lenient overrides on recidivism rates within 3 years of supervision start. The dependent variable in all specifications is a binary variable indicating whether a new felony or misdemeanor was committed within three years of the start of supervision. The key independent variables are "Harsh Override" and "Lenient Override," which are binary variables indicating whether the parole decision was a harsh or lenient override of the initial recommendation. Each column represents a different specification, with additional control variables added sequentially: Demographics (column 2), Criminal History (column 3), Conditions of Supervision (column 4), and Supervision Activities (column 5). Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

As indicated by the coefficients, there is a pronounced differential effect between harsh and lenient overrides. Harsh overrides are consistently associated with a statistically significant reduction in the recidivism rate, with coefficients ranging from -0.0519 to -0.0800 across the different models. This suggests that when parole officers opt for stricter supervision than the algorithm originally suggested, the recidivism rates reduce by approximately 5.2 to 8.0 percentage points. Contrarily, lenient overrides show an initially positive but statistically insignificant relationship with recidivism rates, becoming slightly negative in the models with more controls, albeit still not statistically significant. This indicates that the leniency of parole officers in overriding algorithmic suggestions does not substantially or consistently impact the recidivism rate.

The primary concern in analyzing the effects of override decisions is the potential for selection bias. The discretionary nature of these decisions implies that they are not made randomly. Parole officers, based on their professional experience and interactions, possess insights into an offender's character, family support, and other unobserved determinants that may influence their likelihood to re-offend. Such information, often unaccounted for in standard algorithmic assessments, can influence their decisions to override, leading to non-random assignment in treatment and control groups.

To address the issue of potential selection biases, our study utilized a detailed balance test, with the findings presented in Figure 2. This test scrutinizes the observable attributes near the threshold risk score, adjusting for the level of supervision. It compares parolees who are subject to overrides with those who follow algorithmic recommendations. The results indicate that parole officers often override these recommendations, with harsh overrides correlating with an individual's prior criminal history and the nature of their offenses. This suggests that parole officers might be employing stricter supervision for those they assess as more likely to re-offend, based on their discretion and experience. Hence, we would expect that when they opt for stricter (harsh) overrides, it is for a good reason - i.e., they have an upward bias, predicting a higher chance of reoffending than the algorithm does.

However, when we run your regression analysis, we find that these harsh overrides are associated with a negative coefficient, meaning the actual reoffending rates are lower than expected, even with the parole officers' upward bias. This suggests that the true causal effect of these overrides might be even more negative than the data shows because you'd initially expect the bias to push results in the opposite (positive) direction.

In conclusion, our results underscore the role parole officers play in the parole process, particularly when they opt for harsher supervision levels than suggested by the algorithm. This differential effect highlights the importance of discretion in parole officers' roles and points to potential avenues for improving predictive algorithms by incorporating aspects of professional judgment. At the same time, our findings encourage careful consideration of the potential trade-offs involved when implementing and interpreting algorithmic recommendations in this context.

5 Discussion and Conclusion

The results of our analysis provide valuable insights into the interplay between algorithmic risk scores, parole officer discretion, and recidivism rates. Notably, our findings reveal the potential efficacy of parole officer discretion in reducing recidivism rates, particularly when they opt for harsher supervision than that suggested by the algorithm.

The study embarked on a meticulous balance test to scrutinize the contrasts between parolees affected by harsh overrides from parole officers versus those who adhered to algorithmic guidance. The findings disclosed substantial statistical differences, particularly in variables like 'Violent Non-Sexual Offense', 'Drug Offense', 'Prior Violent Convictions', and conditions such as 'No Victim Contact', 'Electronic Monitoring', or 'Restitution', as well as gender. These significant variances suggest a discernible selection bias in imposing stringent supervisory conditions. Such observations infer that parole officers might be intentionally selecting specific parolee profiles for enhanced supervision, thus modifying the inputs into the parole supervision production function. This targeted approach by parole officers potentially reflects an adjustment of supervision strategies to address perceived risks, an action that recalibrates the standard protocol provided by predictive algorithms.

Contrary to expectations derived from the balance test, the analysis of recidivism rates unveiled a noteworthy trend: individuals subject to harsh overrides exhibited lower rates of recidivism. This counter-intuitive finding suggests that parole officers, leveraging their professional discretion, effectively pinpoint individuals who benefit from stricter supervisory regimes despite certain risk factors. The resulting lower recidivism rates signify that the additional measures associated with harsh overrides are conducive to achieving the desired outcome, namely, the reduction of re-offending.

This phenomenon can be conceptualized as an enhancement of the supervision production function's efficiency. Parole officers, through a combination of algorithmic data and personal discernment, are deploying more rigorous supervision for those they assess as high-risk, leading to an optimization of outcomes. This finding is in alignment with the theoretical expectations of a production function—where a heightened intensity and quality of an input, in this case, supervisory oversight, correlate with improved output, reflected here as lower recidivism rates.

Despite the anticipation that lenient overrides might lead to higher recidivism due to reduced supervisory constraints, the subsequent analysis did not show a statistically significant impact on recidivism rates. This unexpected result indicates that leniency in parole supervision does not necessarily compromise the effectiveness of the intervention. It suggests that the parole officers' discretion, even when resulting in reduced supervision, does not adversely affect recidivism outcomes.

Economically, this outcome can be interpreted as evidence of an efficient allocation of supervisory resources. It implies that parole officers can safely extend leniency without incurring the cost of increased re-offending, thereby conserving resources for cases that necessitate stricter supervision. This efficient resource allocation aligns with economic principles within the production function framework, where the level of input—in this case, the intensity of supervision—is optimized to achieve the most favorable output, which is maintaining low recidivism rates despite a lenient approach.

Our analysis dovetails with the extant literature that highlights the potential advantages of discretion in criminal justice decisions. As researchers such as (Kleinberg et al., 2018) and (Armstrong and Clear, 2017) have noted, there is significant potential value in the human element of discretion when it comes to justice-related decisions. The intuitive judgment of seasoned officers can take into account subtle nuances and specific circumstances that may not be captured by risk assessment algorithms. Our results align with these arguments, demonstrating that override decisions, especially those that impose harsher supervision, can significantly reduce recidivism rates.

Returning to our core findings, we did not identify a significant impact of lenient overrides on recidivism rates, suggesting the nuances of discretion matter. This mirrors literature findings on leniency in criminal justice. Scholars like (Kuziemko, 2013) argue that leniency, while valuable in specific contexts, doesn't always yield better outcomes for offenders. Our results empirically back this perspective.

Building upon this, our results should not be perceived as a blanket endorsement of strict supervision. Parole's aim isn't solely recidivism prevention but also to support successful societal re-entry. Overzealous supervision might counteract this goal, as it could impose undue constraints on parolees, potentially stymying their reintegration (Phelps, 2017).

However, interpreting our findings warrants caution. A potential explanation might be that parole officers, with their depth of experience and nuanced understanding of individual cases, are adeptly pinpointing cases where stricter supervision is merited. On the flip side, when lenient overrides are chosen, they might be misjudging or other uncaptured factors in our study might negate the positive effects of leniency.

Our study also bolsters the discourse on algorithmic predictions in criminal justice. While risk assessment algorithms hold promise in refining processes and curtailing subjective bias, our findings stress the importance of harmonizing these tools with the discretion and expertise of human officers.

Additionally, our research paves the way for further exploration. Future endeavors might spotlight the specifics that make harsh overrides effective in slashing recidivism rates. Delving into conditions where lenient overrides prove advantageous would also be enlightening. Augmenting risk assessment algorithms with these insights might enhance recommendation accuracy and parole supervision efficacy.

To sum up, our research shines a spotlight on the intricate dance between algorithmic recommendations, officer discretion, and recidivism outcomes in parole decisions. The conclusions reemphasize the need for a careful balance between machine-derived predictions and human judgement in the criminal justice domain. As the digital age advances, it becomes paramount to consistently assess and fine-tune the role of algorithms in pivotal societal decision-making spheres, especially in areas as significant as criminal justice.

References

Angelova, Victoria, Will S. Dobbie, and Crystal Yang, "Algorithmic Recommendations and Human Discretion," 2023, (31747).

Armstrong, Todd and Todd R. Clear, *The Role of Parole Officers in the Reentry of Parolees*, Routledge, 2017.

Berk, Richard, "An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism," *Journal of the American Statistical Association*, 2017, *112* (518), 750–765.

_ **et al.**, "Criminal Justice Forecasts of Risk: A Machine Learning Approach," *Springer*, 2018.

_ **, Hoda Heidari, Samira Jabbari, Michael Kearns, and Aaron Roth**, "Fairness in Criminal Justice Risk Assessments: The State of the Art," *Sociological Methods & Research*, 2018, p. 0049124118782533.

Dietvorst, Berkeley J, Joseph P Simmons, and Cade Massey, "Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them," *Management Science*, 2015, *64* (3), 1155–1170.

Dressel, Julia and Hany Farid, "The accuracy, fairness, and limits of predicting recidivism," *Science Advances*, January 2018, *4* (1).

Harcourt, Bernard E., *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*, University of Chicago Press, 2007.

Hawken, Angela and Mark Kleiman, "Managing Drug Involved Probationers with Swift and Certain Sanctions: Evaluating Hawaii's HOPE," 2009.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human decisions and machine predictions," *The quarterly journal of economics*, 2018, *133* (1), 237–293.

Kuziemko, Ilyana, "How should inmates be released from prison? An assessment of parole versus fixed sentence regimes," *The Quarterly Journal of Economics*, 2013, *128* (1), 371–424.

Monahan, John and Jennifer L Skeem, "Risk Assessment in Criminal Sentencing," *Annual Review of Clinical Psychology*, 2016, *12*, 489–513.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, October 2019, *366* (6464), 447–453.

Paparozzi, Mario A and Paul Gendreau, "Parole officer role and discretion in the parole process," *Justice Quarterly*, 2005, *22* (4), 479–508.

Phelps, Michelle S., "The Paradox of Probation: Community Supervision in the Age of Mass Incarceration," *Law & Policy*, 2017, *35* (1-2), 51–80.

Piehl, Anne Morrison and Stefan F. LoBuglio, "DOES SUPERVISION MATTER?," in "Prisoner Reentry and Crime in America," Cambridge University Press, August 2005, pp. 105–138.

Rose, Evan K, "Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders," *The Quarterly Journal of Economics*, 2021, *136* (2), 1199–1253.

Sakoda, Ryan T., "The Architecture of Discretion: Implications of the Structure of Sanctions for Racial Disparities, Severity, and Net Widening," *Northwestern University Law Review*, 2023, *117*, 1213–1276.

Stevenson, Megan T., "Assessing Risk Assessment in Action," *SSRN Electronic Journal*, 2017.

_ **and Jennifer L. Doleac**, "Algorithmic Risk Assessment in the Hands of Humans," *SSRN*, 2021.

Taxman, Faye S, "Supervision–Exploring the dimensions of effectiveness," *Federal Probation*, 2002, *66* (2), 14.

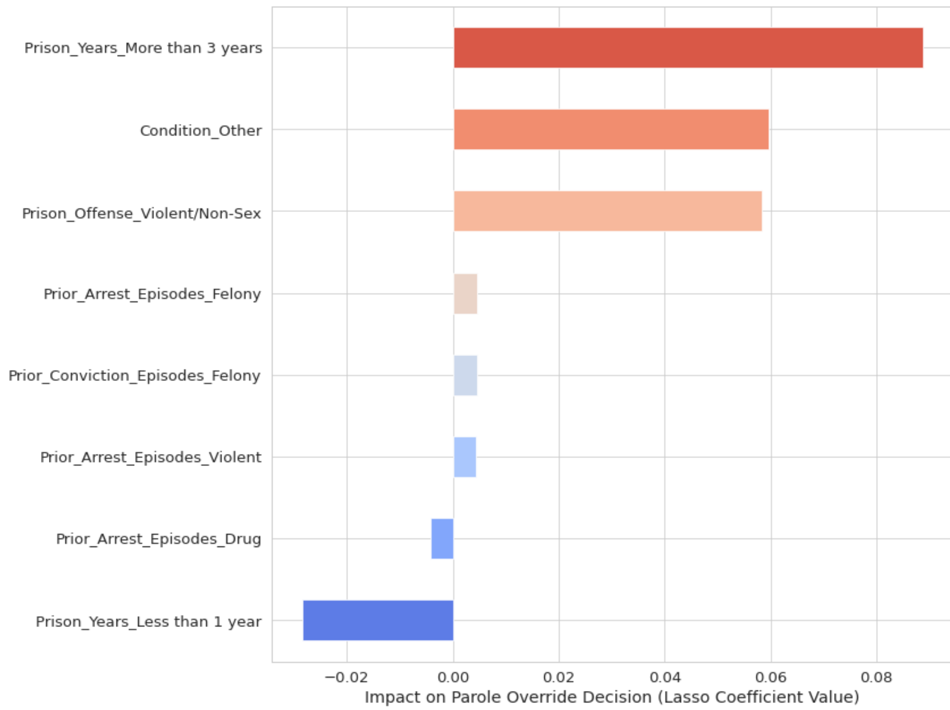
A Appendix Tables and Figures

Appendix Table A1. Summary Statistics of Parole Case Characteristics by Supervision Decision

	All Cases	Follow Algorithm	Lenient Override	Harsh Override
A.Demographics				
Male	0.87	0.87	0.86	0.93
White	0.42	0.44	0.42	0.35
B.Prior Criminal History				
Prior_Arrest_Property	2.22	2.23	2.49	1.89
Prior_Arrest_Drug	1.79	1.85	1.99	1.20
Prior_Arrest_PPViolationCharges	2.31	2.32	2.62	1.93
Prior_Arrest_DVCharges	0.17	0.16	0.17	0.19
Prior_Arrest_GunCharges	0.27	0.26	0.28	0.28
Prior_Conviction_Felony	1.39	1.38	1.46	1.37
Prior_Conviction_Misd	1.74	1.76	1.78	1.59
Prior_Conviction_Viol	0.33	0.30	0.33	0.51
Prior_Conviction_Prop	1.11	1.12	1.28	0.93
Prior_Conviction_Drug	0.77	0.80	0.85	0.50
Prior_Conviction_PPViolationCharges	0.32	0.32	0.33	0.31
Prior_Conviction_DomesticViolenceCharges	0.08	0.08	0.07	0.09
Prior_Conviction_GunCharges	0.13	0.13	0.17	0.15
C.Algorithmic Inputs				
Age_at_Release	34.15	34.31	32.08	35.33
Prison_Years	1.79	1.67	1.86	2.46
Prison_Offense_Property	0.33	0.34	0.37	0.21
Prior_Arrest_Felony	5.70	5.69	6.28	5.21
Prior_Arrest_Misd	3.31	3.35	3.36	2.99
Prior_Arrest_Violent	1.01	0.95	0.98	1.40
Prior_Revocations_Parole	0.09	0.08	0.15	0.09
Prior_Revocations_Probation	0.14	0.14	0.18	0.10
D.Recidivism Outcomes				
Recidivism_Within_3years	0.58	0.59	0.63	0.52
Recidivism_Arrest_Year1	0.30	0.31	0.32	0.25
Recidivism_Arrest_Year2	0.18	0.18	0.20	0.18
Recidivism_Arrest_Year3	0.10	0.10	0.11	0.09
Cases	16140	12332	1959	1849

Notes: This table presents summary statistics for parole case characteristics according to the decisions made (either to follow the algorithm, make a lenient override, or a harsh override). The unit of observation is the individual parole case. The sample includes 16,140 cases from Georgia prisons, with individuals released on discretionary parole to the Georgia Department of Community Supervision (DCS) between January 1, 2013, and December 31, 2015. The categories include demographic details, prior criminal history, algorithmic inputs, and recidivism outcomes for each decision category. All values reported in this table are mean values for the variables indicated in rows.

Appendix Figure A1. Bar Chart of Feature Importance on Parole Override Decisions Using Lasso Regression



Notes: The bar chart above displays the importance of individual features on parole override decisions using the Lasso regression model. Features are ranked by their coefficients, reflecting their contribution to the model. A positive coefficient indicates a feature that pushes the model's prediction higher, while a negative coefficient indicates the opposite. This chart aids in understanding the relative importance and directionality of the features influencing parole override decisions.

Appendix Table A2. Balance Test Results: Observable Attributes at the Risk Score Threshold

	(1) Unconditional	(2) Division FE
Dependent Variable: Risk Score Thresholds		
Male	0.020	0.022
	(0.017)	(0.017)
White	-0.018	-0.018
	(0.012)	(0.013)
Prior_Arrest_Episodes_Violent	-0.009	-0.010
	(0.007)	(0.007)
Prior_Arrest_Episodes_Property	0.002	0.002
	(0.005)	(0.005)
Prior_Arrest_Episodes_Drug	-0.004	-0.005
	(0.005)	(0.005)
Prior_Arrest_Episodes_PPViolationCharges	-0.004	-0.003
	(0.004)	(0.004)
Prior_Arrest_Episodes_DVCharges	-0.011	-0.011
	(0.018)	(0.019)
Prior_Arrest_Episodes_GunCharges	0.011	0.011
	(0.015)	(0.015)
Prior_Conviction_Episodes_Viol	-0.000	-0.002
	(0.015)	(0.015)
Prior_Conviction_Episodes_Prop	0.005	0.004
	(0.008)	(0.008)
Prior_Conviction_Episodes_Drug	0.010	0.011
	(0.010)	(0.010)
Prior_Conviction_Episodes_PPViolationCharges	0.016	0.016
	(0.014)	(0.014)
Prior_Conviction_Episodes_DomesticViolenceCharges	0.002	0.000
	(0.024)	(0.024)
Prior_Conviction_Episodes_GunCharges	0.005	0.004
	(0.019)	(0.019)
Observations	8,352	8,352

Notes: This table displays the results of our balance test, which compares parolees immediately above and below the risk score thresholds that demarcate distinct supervision levels, as visualized in Figure 1. The intent of this analysis is to validate that there is statistical parity in observable attributes for parolees on either side of these thresholds. Each column delineates a unique specification: the initial column illustrates findings from the raw threshold, devoid of control variables, whereas the subsequent column integrates a division fixed effect. The variables used in the regression encompass both demographic attributes and prior criminal records. The main objective is to confirm that variances in outcomes aren't attributed to pre-existing differences in these baseline attributes among the compared groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.