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

Anuar Assamidanov* Nicholas Powell†

October 2023

Abstract

In this paper, we conduct an in-depth examination of 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 a notable decrease in recidivism when officers override algorithmic suggestions. Intriguingly, harsher override decisions consistently lead to reduced recidivism rates, whereas more lenient decisions don't display a significant effect. These findings highlight the crucial role of human discretion in algorithm-based decision-making and provide important insights into potential improvements for predictive algorithms. 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.

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, with the belief that human oversight can provide valuable insights and rectify algorithmic inaccuracies (Dressel and Farid, 2018).

This study explores the often 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 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 (Stevenson 2018; Kleinberg, et al. 2018; Stevenson and Doleac 2022; Albright 2023; Angelova, Dobbie, and Yang 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; Wilson, Naro, and Austin 2007; Hawken and Kleiman 2009; Barnes et al. 2012; Duriez, Cullen, and Manchak 2014; Rose 2021; Sakoda 2023). While each of these areas 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. In this study, I seek to contribute 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 shed light on 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 of 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 choose to lower the supervision level, contrary to the algorithm's recommendation, we did not find any significant effect on recidivism rates. This indicates that reduced supervision in these instances doesn't necessarily translate to a higher likelihood of reoffending.

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 have the potential to 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 parolee must report to their parole officer. Next, there are specific conditions that parolees must comply with during their parole duration. 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 to the parolee. This officer plays a vital role, ensuring compliance with the preceding components and wielding considerable discretion in shaping the parolee-officer relationship.

Parole officers bear the critical task of ascertaining the suitable level of supervision for each parolee, ranging from intensive monitoring to more lenient oversight. Ideally, this supervision level mirrors the parolee's risk of reoffending 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 the parolee's likelihood of reoffending (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 the parolee's unique circumstances (Dietvorst et al., 2015). This intersection of algorithmic recommendations and human discretion forms the crux of our current investigation.

The parole supervision process is further complicated by the fact that parolees are not randomly assigned to parole officers. This non-random assignment can introduce selection bias, as parole officers' supervision styles and override tendencies can influence the outcomes of their assigned parolees (Berk, 2017). Our study proposes the adoption of a randomized assignment of parole officers to parolees, aiming to provide a more robust foundation for analyzing human discretion's effect on parole decisions.

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 takes place 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, encompassing 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 individuals' 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 were 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 was 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 1 presents a detailed examination of our evaluation sample, classified according to the algorithm's predetermined risk score recommendation for the supervision level. Panel A displays the demographic makeup of the sample. Males make up the majority of all categories, accounting for 87% of all cases as per Column 1. The percentage of males is the highest in the group where parole officers decide to override the algorithm's recommendation for a lower supervision level with a more stringent one (93%). Conversely, white individuals are the most numerous in the group where the algorithm's recommendation for a high supervision level is adhered to (44%). They are, however, the least represented in the group where the algorithm's recommendation for a lower supervision level is overridden with a higher one (35%).

Panel B explores the subjects' previous criminal history. The algorithm tends to assign a high-risk score, hence recommending higher supervision levels, to individuals with a history of felony, property, and drug-related arrests. Moreover, these high-risk individuals are more likely to have previous convictions associated with property and drug offenses, as well as parole/probation violation charges. Notably, lenient overrides among high-risk cases and harsh overrides among low-risk cases don't seem random. High-risk individuals who receive a lenient override, for instance, have fewer prior arrests and convictions, are less likely to be on parole or probation, and are more likely to be charged with drug offenses.

Panel C discusses the factors that the algorithm considers. High-risk individuals recommended for a high supervision level are generally younger upon release (average age of 32.08 years) and have spent fewer years in prison (average of 1.86 years). In contrast, low-risk individuals who receive a lenient override, resulting in a lower supervision level, are older upon release (average age of 35.33 years), but have more years of incarceration (average of 2.46 years).

Lastly, Panel D shines a light on recidivism outcomes. Individuals who are released with a higher supervision level, contrary to the algorithm's recommendation, are more likely to re-offend within 3 years (63%) and also have higher rates of arrest in the first year post-release (32%). This stands in stark contrast to individuals who are released with a lower supervision level than the algorithm's recommendation, as they show lower recidivism rates within 3 years (52%) and also lower rates of arrest in the first year (25%).

Table 1. 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.

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, when parole officers exercise discretion and override the recommendation, we notice discernible differences. 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. Descriptive Statistics

		Recidivism Arrest			
		Within_3years	Year1	Year2	Year3
		(1)	(2)	(3)	(4)
Recommend	Discretion				
High	Follow Algorithm	0.640	0.340	0.193	0.108
	Lenient Override	0.615	0.311	0.193	0.111
	Harsh Override	0.574	0.282	0.197	0.094
Specialized	Follow Algorithm	0.693	0.390	0.201	0.102
	Lenient Override	0.670	0.358	0.211	0.100
Standard	Follow Algorithm	0.479	0.234	0.154	0.091
	Harsh Override	0.467	0.215	0.164	0.088

Notes: This table displays the recidivism arrest rates at different time intervals (within 3 years, Year 1, Year 2, and Year 3) for parolees in different supervision levels (High, Specialized, Standard) and decision categories (follow algorithm, harsh override, lenient override). All reported values in this table are averages. The term 'Follow Algorithm' refers to cases where parole officers followed the recommendations of the predictive algorithm. 'Harsh Override' refers to cases where the parole officers decided for a stricter supervision level than recommended by the algorithm, while 'Lenient Override' refers to cases where officers assigned a more lenient supervision level than the algorithm's recommendation. This analysis helps to understand the impact of different types of supervision decisions on the recidivism arrest rates.

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. I am 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.

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:

$$Recidivism_i = \alpha_0 + \alpha_1 Override_i + \alpha_2 Threshold_i + \alpha_3 X_i + \nu_i \tag{1}$$

In this equation, $Recidivism_i$ represents the propensity of individual i to reoffend. The term $Override_i$ captures the parole officers' decision regarding individual i—whether to follow the algorithmic recommendations or override them. Meanwhile, $Threshold_i$ signifies the risk score that indicates potential shifts in supervision levels. The vector X_i encompasses a range of control variables, potentially covering demographic factors, past criminal history, attributes of the parole officer, and other pertinent covariates. The term ν_i represents the error term tied to individual i. The coefficient of primary interest in this equation, α_1, encapsulates the causal influence of override decisions on recidivism.

Diving deeper into the nature of the override decisions, we wanted to discern the distinct impacts of 'harsh' and 'lenient' overrides. To achieve this granularity, our methodology introduces interaction terms into the analysis. The enhanced model can be described as:

$$Recidivism_i = \gamma_0 + \gamma_1 O_i + \gamma_2 O_i \times Harsh_i + \gamma_3 O_i \times Lenient_i + \gamma_4 Threshold_i + \gamma_5 X_i + \zeta_i \tag{2}$$

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.

We further augmented our analysis by integrating an identification strategy akin to the Difference in Differences (DID) method with dummy outcomes. This approach primarily concentrated on the interaction between override decisions and the levels of supervision. For elucidation, we drew parallels with the study by Price (2010) which investigated racial biases among NBA referees. Analogously, we explored differences in the probability of recidivism under high and specialized supervision categories conditioned on whether the algorithm's recommendation was followed or overridden with a harsh decision. The model encapsulating this approach is:

$$Recidivism_i = \beta_0 + \beta_1 Harsh_i + \beta_2 Specialized_i + \beta_3 Harsh_i \times Specialized_i + \gamma_0 X_i + \zeta_i \tag{3}$$

In the presented model, the dependent variable, $Recidivism_i$, delineates the propensity of an individual, represented by i, to reoffend. This model seeks to ascertain the interplay between varying levels of supervision and the impact of harsh override decisions on this propensity. The term $\beta_1 Harsh_i$ signifies the marginal impact of a harsh override on recidivism when specialized supervision is not a factor. On the other hand, $\beta_2 Specialized_i$ measures the effect of specialized supervision on recidivism without considering the type of override. Central to the Difference in Differences (DID) design is the interaction term $\beta_3 Harsh_i \times Specialized_i$. This coefficient quantifies the additional effect of a harsh override specifically for individuals subjected to specialized supervisionn, compared to those under high supervision level. If β_3 is statistically significant and positive, it infers that a harsh override decision amplifies the likelihood of reoffending for those under specialized supervision, relative to their counterparts high supervision level. The term $\gamma_0 X_i$ incorporates control variables pertinent to recidivism, while ζ_i denotes the error term encapsulating unobserved determinants.

$$Recidivism_i = \beta_0 + \beta_1 Lenient_i + \beta_2 Standard_i + \beta_3 Lenient_i \times Standard_i + \gamma_0 X_i + \zeta_i \quad (4)$$

This model parallels on harsh overrides, but pivots to focus on lenient overrides and their corresponding supervision levels. As illustrated in Figure 1, the DID framework provides a detailed view of how the effects unfold. It dissects the relative effect of a harsh or leninent override across different supervision levels, offering nuanced insights into the relationship between discretion, supervision level, and the probability of recidivism.

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

Building upon the analytical framework from the previous section, this results segment delves into our key findings. We start with the primary equation anchored on the threshold assumption and incorporate the parole officers' override decisions. As outlined in our methodology, we then estimate the causal effect of these overrides on recidivism rates, setting the stage for the subsequent detailed analysis.

Turning to equation 1, 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

	(1)	(2)	(3)	(4)	(5)
Outcome: New Felony/Mis within 3 Years of Supervision Start					
Override	-0.0231**	-0.0281***	-0.0281***	-0.0329***	-0.0271***
	(0.0093)	(0.0092)	(0.0091)	(0.0091)	(0.0090)
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: Table presents the estimated effects of override decisions 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 variable is "Override," which is a binary variable indicating whether the parole decision overridden 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The coefficients for the override variable across all models are negative and statistically significant, suggesting that an override decision, whether harsh or lenient, consistently reduces the probability of new felony/misdemeanor within 3 years of supervision start. The magnitudes of the coefficients range from -0.0231 to -0.0329, indicating a reduction in recidivism rates of approximately 2.3 to 3.3 percentage points due to an override. Interestingly, this impact remains robust to the addition of further control variables, such as criminal history, conditions of supervision, and supervision activities. This consistency speaks to the robustness of our results and suggests that overrides - decisions by parole officers to deviate from the algorithmic recommendation - may indeed have a beneficial impact on reducing recidivism rates. This evidence, while tentative, invites further investigation into the mechanisms and nuances behind such override decisions.

Expanding our analytical framework, we aimed to explore further the differential effects of harsh and lenient override decisions on recidivism rates by incorporating interaction terms with the override decision variable. Table 4 provides a comprehensive overview of our findings based on equation 2.

Table 4. 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 what was originally suggested by the algorithm, 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 have a substantial or consistent impact on 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 mitigate the concerns arising from potential selection biases, we rigorously implement two pivotal strategies. Initially, we conduct a meticulous balance test, the results of which are presented in Table A1 and are further elucidated in the supplementary appendix. This test ascertains that the observable attributes near the threshold are statistically indistinguishable after controlling for the underlying risk score. Such a verification provides strong evidence indicating the absence of any notable differences, in the context of observed characteristics, between parolees who experience overrides and those who adhere to the algorithmic recommendations. This foundational step assures the robustness of our subsequent analyses and findings, ensuring that our results are not confounded by unobservable differences at the threshold.

Additionally, the algorithm predicts a certain level of parole supervision based on various factors. Parole officers, however, sometimes override these recommendations, often thinking stricter supervision is needed for those parolees they consider more likely to re-offend. Given their discretion and experience, 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 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.

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 judgement 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.

In Figure A1, we provide a detailed exploration of the intricacies of override decision-making in the parole system using the Lasso regression model. The adoption of Lasso was motivated by its capacity for feature selection by down-weighting certain coefficients towards zero, thereby pinpointing critical predictors. By evaluating the non-zero coefficients from the Lasso regression, we discerned the paramount variables influencing these decisions.

The 'Condition_Other' feature emerged as significant with a coefficient of 0.0596. This feature amalgamates a range of parole conditions, such as 'No Victim Contact', and participation in specified programs including 'Electronic Monitoring' and 'Sex Offender Registration/Programs'. The importance assigned to this feature by the Lasso model underscores the escalating relevance of compliance with distinct parole conditions in override decisions. It is intriguing that this focus on certain conditions may supersede the significance of the primary conviction offense. These findings spur deeper introspection into the hierarchy of these parole conditions and their expansive implications in the justice system and the rehabilitation of offenders.

Of equal note is the feature 'Prison_Offense_Violent/Non-Sex', assigned a coefficient of 0.0585, which represents the primary prison convictions for non-sexual violent offenses. Its pronounced weight in the model signals that parole officers may harbor amplified concerns about potential recidivism involving violent crimes. This accentuates the delicate equilibrium parole officers aim to achieve: safeguarding community well-being whilst recognizing the rehabilitation prospects of the inmate.

Furthermore, the time inmates spend in prison also plays a substantial role in the decision-making process. The feature 'Prison_Years_More than 3 years' , having a coefficient of 0.0887, emphasizes that inmates with prolonged incarceration durations might face challenges in securing parole overrides. Conversely, inmates with a shorter duration, as represented by the feature 'Prison_Years_Less than 1 year' with a coefficient of -0.0284, might experience a slightly favorable stance in some circumstances.

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.

While these harsher decisions might require a greater allocation of resources, the trade-off is a marked improvement in recidivism. Meanwhile, when officers opt for leniency, the outcomes in terms of reoffending are indifferent from those of following algorithmic recommendations. What's interesting is that such lenient decisions, while achieving similar outcomes as the algorithm, come at a reduced economic burden — making them a cost-effective choice.

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

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.

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.

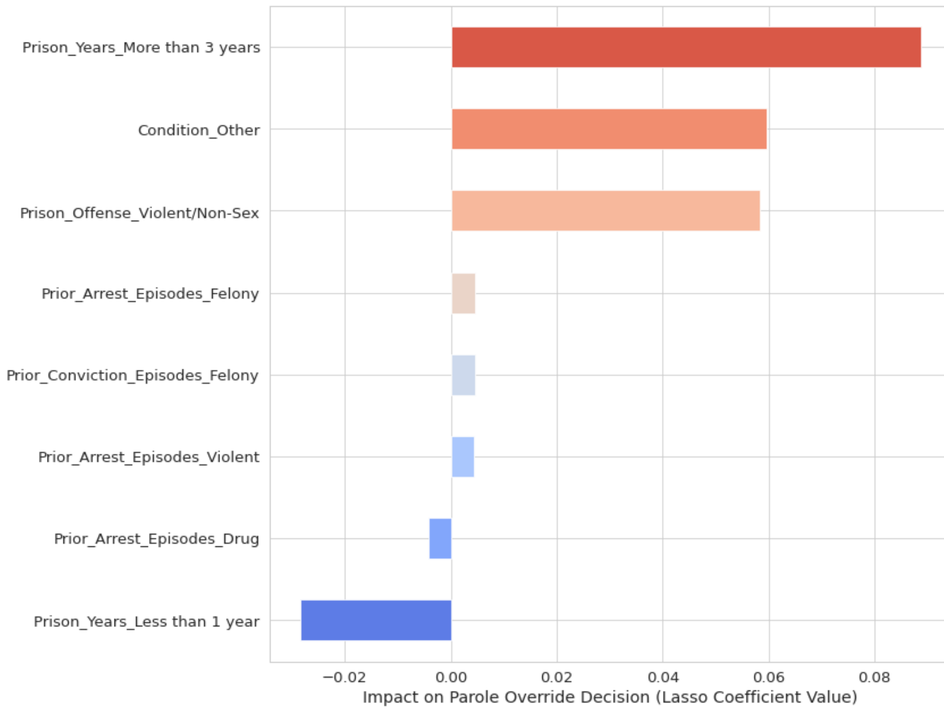
Paparozzi, Mario A and Paul Gendreau, "Parole officer role and discretion in the parole process," *Justice Quarterly*, 2005, *22* (4), 479–508.

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

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

A Appendix Tables and Figures

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 A1. 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.