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

Anuar Assamidanov* Nicholas Powell†

May 2023

Abstract

In this paper, we conduct an in-depth examination of the interplay between predictive algorithms and human discretion in parole supervision decisions. Leveraging a unique methodological framework combining Regression Discontinuity Design (RDD) and random assignment, we explore the impact of parole officers' override decisions on recidivism rates. Our analysis indicates a significant increase in recidivism as parolees transition from standard to high supervision based on algorithmic risk scores alone. However, when parole officers' override decisions are incorporated, harsher decisions consistently result in lower recidivism rates, while lenient decisions have no significant impact. 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, with a particular focus on parole supervision decisions. 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 overcome this selection challenge, I propose the use of a method involving the random assignment of parole officers to parolees. This method can handle the missing data issue and estimate potential outcomes for both the observed and unobserved supervision levels (Berk et al., 2018b). The data used in this study is derived

from persons released from Georgia prisons on discretionary parole to the custody of the Georgia Department of Community Supervision (DCS) between January 1, 2013, and December 31, 2015. It includes over 25,000 de-identified records containing information such as demographics, prison and parole case information, prior community supervision history, conditions of supervision, and recidivism rates. The analysis of this extensive dataset forms the backbone of our study.

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, I found that when parole officers decide to assign a higher supervision level than suggested by the algorithm, recidivism rates slightly 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.

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 officers bear the critical task of ascertaining the suitable level of supervision for each parolee. This level can span from intensive monitoring to more lenient oversight. Ideally, the 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 subsequently predicts the parolee's probability of reoffending (Berk et al., 2018a). Leveraging this risk score, the algorithm proposes a supervision level intended to minimize the recidivism risk while optimizing the allocation of parole supervision resources (Monahan and Skeem, 2016).

Nevertheless, parole officers maintain the discretion to override these algorithmic recommendations. Such overrides may transpire when the officer has access to additional information not considered by the algorithm, or when they believe that the algorithm does not sufficiently comprehend the unique subtleties of the parolee's circumstances (Dietvorst et al., 2015). This intersection of algorithmic recommendations and human discretion is the core of the current investigation.

The parole supervision process is further convoluted by the reality that parolees are not randomly assigned to parole officers. Such non-random assignment may introduce selection bias, as the supervision styles and override propensities of parole officers may potentially affect the outcomes of the parolees under their charge (Berk, 2017). To circumvent this issue, our study advocates for the implementation of a randomized assignment of parole officers to parolees, which would serve as a more robust platform for analyzing the impact of human discretion on parole supervision decisions.

In this study, the algorithm-generated risk score is numerical, ranging from 1 to 10. Each score range corresponds to a recommended supervision level. In detail, a risk score between 1 and 5 implies a standard supervision level—geared towards individuals with a relatively low risk of recidivism. Conversely, a risk score between 6 and 8 necessitates a high supervision level, typically for individuals with a more considerable criminal history or other risk factors. Lastly, a risk score between 9 and 10 calls for a specialized supervision level, intended for individuals with a serious criminal history or high-risk factors.

While these risk scores and corresponding supervision levels function as guiding

principles, it is pivotal to remember that parole officers maintain the right to override the algorithm's recommendations. This discretionary override is based on their professional judgment and any additional, possibly uncaptured, information about the parolee's situation. The magnitude and implications of these discretionary overrides serve as the foundation of our 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 detailed overview of the recidivism arrest rates segmented by the type of recommendation and discretion applied in parole cases. Specifically focusing on the high supervision level, it is noteworthy that parolees subject to the "Follow Algorithm" have the highest incidence of recidivism arrests within a three-year period post-release, pegged at 64%. Furthermore, this group manifested a recidivism arrest rate of 34% in the first year alone. Contrastingly, the rates for the "Harsh Override" and "Lenient Override" were marginally lower during the same period, underscoring the nuanced impact of these different approaches within the high supervision level.

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
	Harsh Override	0.615	0.311	0.193	0.111
	Lenient Override	0.574	0.282	0.197	0.094
Specialized	Follow Algorithm	0.693	0.390	0.201	0.102
	Harsh Override	0.670	0.358	0.211	0.100
Standard	Follow Algorithm	0.479	0.234	0.154	0.091
	Lenient 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.

Turning to the specialized supervision level, an interesting pattern is observed. Parolees under the "Follow Algorithm" again showed the highest rate of recidivism arrests within three years at 69.3%. However, the recidivism rates for the "Harsh Override" recommendation were quite close, marking a significant increase compared to the high-risk category.

In the standard supervision level, there is a considerable drop in recidivism rates compared to the other groups. Parolees with a "Follow Algorithm" in this category show a recidivism arrest rate of 47.9% within three years. The "Lenient Override" in this category leads to the lowest recidivism rate overall.

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, I adopt a robust methodological approach known as Regression Discontinuity Design (RDD). This methodology is uniquely suitable for our research, given its reliance on arbitrary cutoff points or thresholds. In the context of our study, these thresholds are represented by the risk scores that trigger a shift in the level of supervision. For example, a parolee may be transitioned from standard supervision to high supervision, or from high to specialized supervision, based on their risk score crossing a certain predetermined threshold.

The underlying assumption of RDD, which also forms the basis of its strength, is that around these cutoff points, the assignment of individuals to different supervision levels can be treated as essentially random. If we observe any significant discontinuity or 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.

In the Regression Discontinuity Design (RDD) framework, we first define a running variable r_i, which in this case is the risk score of individual i. The cutoff value c

represents the risk score threshold at which the level of supervision changes. We denote the level of supervision as D_i, and it is equal to 1 if the risk score r_i is greater than or equal to c, and 0 otherwise.

The balance check in the RDD framework can be assessed by regressing the covariates X_i on the treatment variable D_i and the running variable r_i, while controlling for their interaction:

$$X_i = \beta_0 + \beta_1 D_i + \beta_2 r_i + \beta_3 D_i \cdot r_i + u_i \tag{1}$$

The null hypothesis in this balance check is that $\beta_1 = 0$, which means that there is no discontinuity in the covariates at the cutoff point, implying the groups on either side of the threshold are comparable.

Next, we estimate the local average treatment effect (LATE) of the supervision level on the outcome variable (recidivism) Y_i where I focus on the effect of supervision level on recidivism while excluding override cases. This allows us to discern the impact of supervision level absent the influence of parole officers' overrides, using the following regression model:

$$Y_i = \alpha_0 + \alpha_1 D_i + \alpha_2 r_i + \alpha_3 D_i \cdot r_i + v_i \tag{2}$$

The parameter of interest is α_1, which captures the discontinuity in the recidivism rate at the risk score cutoff c. If α_1 is significantly different from zero, it implies that the supervision level has an impact on recidivism. By estimating this model, we can capture the local average treatment effect (LATE) of the supervision level on recidivism rates, in the absence of parole officer overrides. This provides insights into the inherent effectiveness of the supervision levels as determined by the risk scores.

Our analytical framework expands on traditional methods. Upon identifying a discontinuity at the threshold through Regression Discontinuity Design (RDD), we take advantage of the random assignment of parole officers to parolees. This allows us to account for any officer-specific characteristics that could confound our results.

We then integrate the override decisions into our analysis. This integration aims at investigating the impact of parole officers' override decisions on recidivism rates. This approach enables us to scrutinize the effectiveness of the initial algorithmic recommendations and the consequences of parole officers' discretion when they decide to override these recommendations.

We first perform a balance check for observable covariates. For this, we compare

individuals who had their supervision level overridden with those who did not. Our comparison is made while controlling for the risk score and demographic variables. The equation is:

$$O_i = \beta_0 + \beta_1 RiskScore_i + \beta_2 X_i + \epsilon_i$$

Here, O_i represents the override decision for individual i, $RiskScore_i$ denotes the risk score of individual i, and X_i signifies the demographic variables of the individual i. The objective is to determine whether the β_1 and β_2 coefficients are statistically significant. If they are not, it would suggest that the override and non-override decisions are similar when considering observable covariates.

Next, we estimate the causal effect of override on recidivism rates. For this, we treat the override as the treatment variable and recidivism as the outcome. The regression equation is:

$$Recidivism_i = \alpha_0 + \alpha_1 Override_i + \alpha_2 RiskScore_i + \alpha_3 X_i + \nu_i$$

Here, $Recidivism_i$ is the recidivism outcome for individual i, and $Override_i$ represents the override decision. α_1 is the coefficient of interest that quantifies the impact of override on recidivism.

Finally, to delve deeper into the nature of the decision and whether the effect comes from harsh or lenient overrides, we decompose the effect by introducing interaction terms with the override decision variable:

$$Y_i = \gamma_0 + \gamma_1 O_i + \gamma_2 O_i * Harsh_i + \gamma_3 O_i * Lenient_i + \gamma_4 RiskScore_i + \gamma_5 X_i + \zeta_i$$

In this equation, $Harsh_i$ and $Lenient_i$ are binary variables indicating whether the override was harsh or lenient for individual i. γ_2 and γ_3 measure the differential effect of harsh and lenient overrides on recidivism rates.

In conclusion, this multi-layered methodological approach provides a comprehensive analysis of the impact of parole officers' override decisions on recidivism rates. It allows us to control for a range of factors and to delve deeper into the nuances of these decisions.

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 Regression Discontinuity Design (RDD) method. As illustrated in the table in the appendix, the results of the balance check regression (equation 1) exhibit no discontinuity in observable characteristics at the threshold, which validates our RDD approach.

Table 3 presents the regression results corresponding to equation 2, where we assess the effect of the supervision level on recidivism within a 3-year period, excluding override cases. The table is designed to elaborate on the results obtained from different model specifications (Models 1 through 5), where each model gradually controls for additional covariates.

For the cutoff between risk scores 5 and 6, the coefficient is positive and statistically significant across all specifications. This implies that as parolees move from a risk score of 5 to 6, thereby shifting from standard supervision to high supervision, there is an increase in recidivism rates. This suggests that the elevated supervision level alone, even without considering override decisions, leads to a higher likelihood of recidivism. The magnitude of this effect slightly decreases as we progressively control for demographics, criminal history, conditions of supervision, and supervision activities but remains significant throughout.

Table 3. RDD Estimates: New Felony/Misdemeanor within 3 Years

	(1)	(2)	(3)	(4)	(5)
Outcome: New Felony/Mis within 3 Years of Supervision Start					
Cutoff 5 and 6 Risk Score	0.0772***	0.0729***	0.0642***	0.0629***	0.0583***
	(0.0121)	(0.0121)	(0.0119)	(0.0120)	(0.0117)
Cutoff 8 and 9 Risk Score	-0.0087	-0.0094	-0.0007	-0.0008	0.0056
	(0.0114)	(0.0114)	(0.0111)	(0.0111)	(0.0108)
Demographics		Y	Y	Y	Y
Criminal History			Y	Y	Y
Conditions of Supervision				Y	Y
Supervision Activities					Y
N	6157	6157	6157	6157	6006

Notes: The table presents the results of a Regression Discontinuity Design (RDD) estimate for the occurrence of a new felony or misdemeanor within 3 years of the start of supervision. The outcome is regressed against two risk score cutoff groups (5 and 6, and 8 and 9), with control variables added in each column. Column (1) includes no control variables. Column (2) adds demographic controls. Column (3) includes demographic and criminal history controls. Column (4) adds conditions of supervision to the previous controls. Finally, column (5) includes supervision activities as a control variable in addition to the previous controls. Standard errors are presented in parentheses below the coefficients. N represents the number of observations. The decrease in observations in column (5) indicates missing data for the supervision activities control. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In contrast, the coefficient for the cutoff between risk scores 8 and 9 is not statistically significant in all specifications, which means there is no detectable discontinuity in recidivism rates as parolees transition from high supervision to specialized supervision based solely on the algorithmic risk score. This suggests that, without considering override decisions, moving from a high to specialized supervision level does not significantly alter the recidivism rates.

These findings provide key insights into the effectiveness of supervision levels as determined by the risk scores. However, to obtain a comprehensive understanding of the impact of parole officers' discretionary override decisions on recidivism rates, further analysis including override cases is needed. This will form the basis of our subsequent discussion in the following sections.

Moving forward with our analysis, we incorporate the parole officers' override decisions into the assessment. As described in the methodology section, we first perform a balance check for observable covariates, followed by an estimation of the

causal effect of override on recidivism rates.

The regression results corresponding to the balance check (equation 3), illustrated in the table in the appendix, indicate that observable characteristics between the overridden and non-overridden groups are statistically similar once we control for the risk score and demographic variables. This confirms that the composition of these groups does not significantly differ in observed attributes, reinforcing the credibility of our subsequent estimations.

Turning to equation 4, which estimates the causal effect of override on recidivism, Table 4 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 4. 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 su-

pervision 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 5 provides a comprehensive overview of our findings based on equation 5.

Table 5. 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. 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 our investigation into the intricate dynamics of override decision-making in the parole system, we turned to the XGBoost machine learning model. The choice of XGBoost was driven by its strong predictive capabilities. By applying the SHAP

method, we were able to pinpoint the dominant variables in these decisions.

The "ConditionOther" feature emerged prominently. This feature encompasses various parole conditions, like 'No Victim Contact', and participation in designated programs such as 'Electronic Monitoring' or 'Sex Offender Registration/Programs'. The high emphasis on this feature suggests that compliance with certain parole conditions is increasingly crucial in override decisions. It is noteworthy that this emphasis on specific conditions might even outweigh the importance of the primary conviction offense. Such insights beckon a deeper examination of how these parole conditions are prioritized and their broader role in the justice system and offender rehabilitation.

Another notable feature is "PrisonOffenseViolent/Non-Sex", referring to primary prison convictions for non-sexual violent offenses. Its strong presence indicates that parole boards may exhibit heightened caution with potential recidivism involving violent crimes. This highlights the challenging balance that parole boards strive to strike: ensuring community safety while acknowledging the rehabilitation potential of the offender.

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.

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 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).

Our study 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.

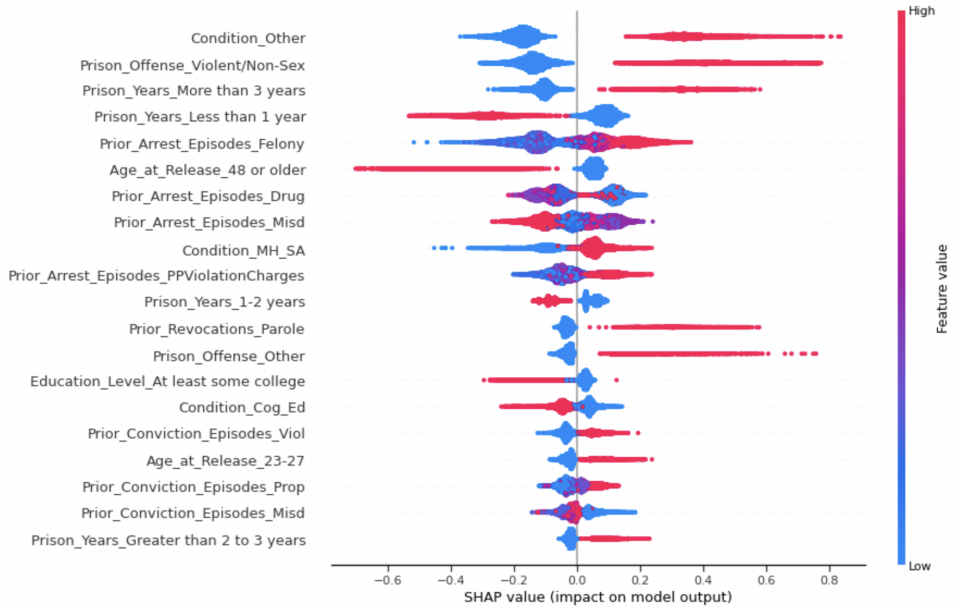
Paparozzi, Mario A and Paul Gendreau, "Parole officer role and discretion in the parole process," *Justice Quarterly*, 2005, *22* (4), 479–508.

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. SHAP Summary Plot of Feature Impact on Parole Override Decisions Using XGBoost Model



Notes: The above SHAP summary plot visualizes the impact of individual features on parole override decisions using the XGBoost model. Each dot represents a sample from our dataset, with the x-axis indicating the SHAP value or impact on the model prediction. Features are ranked by importance from top to bottom. The color spectrum, ranging from blue to red, denotes the feature's value, with blue indicating low and red indicating high values. A dot's horizontal location reflects whether the feature pushes the model's prediction higher (to the right) or lower (to the left). This visualization aids in understanding the relative importance and directionality of the features influencing parole override decisions.