

# How Much do Mandatory Minimums Matter?\*

Spencer Cooper

February 22, 2024

## Abstract

I estimate the causal increase to sentence lengths created by mandatory minimum (MM) eligibility for federal drug crimes. Estimating sentence effects is complicated by case manipulation near the MM eligibility cutoff, which creates a bunching point in case eligibility. I exploit this bunching point to decompose the impacts of MM eligibility into effects driven by the law and effects driven by manipulation. I find that without manipulation, MM eligibility increases sentence length by about 10 months on the intensive margin. This effect is consistent across racial groups with small heterogeneity by drug type and by criminal history. However, through case manipulation, I find that minority defendants receive an additional 5-7 months to their sentence lengths, while manipulation has no significant impact on White sentence lengths. Using a unique data set of US Attorneys, I show that minorities are also disproportionately bunched on the extensive margin. These results imply that race disparities in MM eligibility are driven by legal actors' decisions to manipulate cases.

---

\*Affiliation: University of Connecticut, Department of Economics. Email: [spencer.cooper@uconn.edu](mailto:spencer.cooper@uconn.edu). I thank Evan Taylor for guidance and support throughout this project. I also thank Katherine Barnes, Tiemen Wouterson, Daniel Herbst, Hidehiko Ichimura, Jason Kreag, Stephen Ross, David Simon, and several anonymous assistant US Attorneys and Assistant US Attorneys for helpful comments and insight.

Mandatory minimum (MM) sentencing is one of the most controversial practices in the United States criminal justice system. Mandatory minimums restrict a judge’s ability to give a sentence length below a specified amount. Thus, MM charges are expected to increase sentence lengths for eligible cases. This practice applies to several crime types in the federal system, including weapon crimes, sex crimes, certain economic crimes, and most commonly, drug crimes. Federal drug trafficking offenses may be eligible for a 5-year or 10-year MM sentence depending on the drug type and the charged quantity of drugs. Many cases are affected by this practice; in 2019 when looking at the five most common drug types, over 86% of cases were eligible for a MM sentence, with over 69% of cases meeting the threshold for a 10-year MM charge.

The controversy surrounding mandatory minimums can be broadly summarized in two points. First, MM eligibility may increase sentence lengths for non-violent or low-criminal history offenders, who may be inefficiently punished due to the constraints created by minimum sentencing. Second, MM eligibility is believed to disproportionately impact certain drug types and racial groups. Because of low weight thresholds, MM sentencing raises concerns of disproportionate punishment for crack cases compared to other drug types. In conjunction with this, MM sentencing is often assumed to disproportionately punish minority race defendants, especially Black defendants, compared to White counterparts.<sup>1</sup> In this paper, I investigate both of these concerns.

Estimating the impacts mandatory minimums have on sentence lengths is complicated by several factors. One issue is that MM charges are not always binding and many cases that are eligible are not formally charged with the MM statute, a fact that has been identified and discussed in prior literature (Rehavi and Starr 2014; Bjerk 2017b). Non-binding sentences can occur for several reasons, but are largely driven by plea negotiation mechanisms. For this reason I consider the impact of eligibility rather than if a case carries a MM charge at sentencing. Eligibility captures the impact of a binding MM charge as well as cases that are impacted by the law only through changes in bargaining power. Thus, by measuring eligibility impacts, I capture the effect for all cases, not just those that appear binding or receive formal MM charges at sentencing.

However, the primary challenge in estimating the causal impact of MM laws on sentence length come from selection issues. For most cases, MM eligibility is determined by the drug quantity listed at sentencing, measured by weight. If the weight at sentencing is greater than some threshold value, an individual is eligible to be charged with a 5 or 10-year mandatory minimum. However, this quantity does not need to equal the weight recorded at arrest. Essentially, this means that cases near the MM threshold may have their charged quantity manipulated in such a way that their eligibility is determined in part by legal actors. Tuttle (2023) and Cooper (2023) note the

---

<sup>1</sup>Rehavi and Starr (2014) attributes significant disparities in sentence lengths to MM charges.

presence of significant bunching in the number of cases right above the 10-year MM threshold weight, evidence of this manipulation. This means that certain cases are systematically moved from 5-year MM eligibility to 10-year MM eligibility, thereby distorting the relationship between charged weight and sentence length.

I exploit this bunching point to estimate the causal impact of MM eligibility on federal drug trafficking sentence lengths. I then specifically consider how MM effects contribute to race gaps in sentencing on both the intensive and extensive margin. To estimate intensive margin effects, I decompose MM eligibility impacts into effects driven by the law and effects driven by legal actor discretion. This is done by separately fitting the distribution of sentence length over drug weight using data only in non-manipulated regions on either side of the 10-year eligibility cutoff. I then extrapolate the fitted distributions into the manipulated regions to create distributions under the counterfactual scenario where there is no manipulation. The discontinuity between these counterfactual distributions at the threshold weight is what I call the legal effect. It is the impact of MM laws for cases without manipulation. The distance between the actual sentence length and the higher counterfactual distribution (the one fit from the right-side of the distribution) is what I call the manipulation effect. It is a combination of two separate effects: (1) selection bias created by manipulation choices, and (2) the impact of being pushed from the 5-year eligibility into 10-year eligibility separate from the legal effect. That is, cases at the bunching point may have high sentence lengths because these cases are fundamentally different than non-manipulated cases, but also because defendants may receive some sentence length premium at the higher eligibility level. I provide evidence that manipulation effects are at least in part driven by sentence length premiums at the higher eligibility level.

I find that, across the five most common drug types, 10-year MM eligibility increases sentence lengths by about 10 months (or 16% above the mean) purely through the legal effect. I find that the manipulation effect for the average case is around 17 months. Further, the manipulation effect persists with the inclusion of controls, and dominates the missing sentence lengths. This indicates that manipulation increases sentence lengths above the legal effect, with a sentence premium estimated to be around 7 months. The results suggest that on average, all cases see an increase to their sentence length if eligible for the 10-year mandatory minimum, but for cases that are manipulated up to the 10-year eligibility, the increase is significantly higher.

The main finding of the intensive margin analysis is that racial disparities are almost entirely driven by manipulation. I find that the legal effect is nearly identical across race groups, and is fairly similar across drug type and criminal history level. However, the manipulation effect is entirely driven by racial minorities, with effects mostly local to cocaine and heroin cases. This

means that the legal impact of MM eligibility has similar effects for most cases, but legal actors use discretion to choose Black and Hispanic individuals to get significantly higher sentence lengths through manipulation. Manipulation effects are also present for low criminal history offenders, suggesting that policies such as the safety valve provision may not adequately shield low level offenders from MM effects. I then present some descriptive statistics to identify which types of cases are being selected for manipulation. I find that cases at the bunching point tend to be the “worst” types of cases with higher propensities for gun involvement, higher probability to go to trial, and higher criminal histories; this is evidence that manipulation is occurring upward.

Next I utilize a unique data set of US Attorney spells to consider race effects on the extensive margin. Comparing the US Attorney bunching propensity measure by race, I find that racial minorities are significantly more likely to be bunched compared to White counterparts. Specifically, a 10 percentage point increase in bunching propensity increases the probability of being bunched for Blacks 3 percentage points more than for Whites. This bunching activity is shown to drive race gaps in sentence lengths, as minority cases adjudicated under high bunching administrations receive significantly higher sentence lengths than White cases, but only at the MM threshold weight. Taken together, these results suggest that race gaps in MM sentencing occur on both the intensive and extensive margin and are driven by legal actor discretion.

This paper contributes to literatures on the welfare impacts of sentencing structure, legal actor discretion, and how each of these may drive racial disparities. Recent literature has suggested prosecutor discretion may actually limit racial disparities, specifically in connection with enforcement decisions (Shaffer 2023; Shaffer and Harrington 2017). Findings from this paper support previous results; that race disparities are largely driven by discretion and are likely tied to prosecutor choices. On a smaller scale, this paper also adds to the drug crime MM sentencing literature. Findings among these studies vary, with some papers downplaying the significance of mandatory minimums (Bjerk 2005; Fischman and Schanzenbach 2012; Bjerk 2017a; Bjerk 2017b) and others finding significant impacts on sentencing generally and in contributing to racial disparities (Rehavi and Starr 2014; Tuttle 2023; Didwania 2020). This paper uses bunching techniques similar to Diamond and Persson (2017), leniency measures similar to Goncalves and Mello (2021), and exploits the same bunching point as seen in Tuttle (2023) and Cooper (2023).

Bjerk (2017b) is the paper most closely related to this work. It provides a descriptive analysis of the sentence length impacts of being convicted at particular charging weights for fiscal years 2011 and 2012. The main analysis shows that MM eligibility rates are similar across drug types, the safety valve provision effectively prevents low-level offenders from receiving large sentence length

increases, and that effects for crack are minimal.<sup>2</sup> I contribute to this paper in four ways: by identifying the causal impact of MM laws on sentence length, by disentangling causal effects into legal and manipulation effects, by assessing the degree to which low criminal history defendants are protected from MM effects, and by giving a more comprehensive view of MM impacts on racial and drug type disparities.

## II Background

### Mandatory Minimum Sentencing of Federal Drug Cases

The main eligibility criteria for MM eligibility is drug quantity. In order for the mandatory minimum to apply, the weight at sentencing for one drug type must meet or exceed the set threshold weight.<sup>3</sup> There are two separate thresholds, one for a 5-year and one for a 10-year mandatory minimum. A small subset of cases are eligible for MM penalty without meeting quantity thresholds. This can occur in two ways, if the crime involved death or serious injury, or if the defendant has committed a serious prior drug offense. Including, excluding, or controlling for this subset does not substantially change results or significance of findings.<sup>4</sup>

Currently, MM sentencing applies to eight illegal or controlled substances: powder cocaine, crack, heroin, marijuana, methamphetamine, lysergic acid diethylamide (LSD), phenylcyclohexyl piperidine (PCP), and fentanyl. Due to few number of cases among the last three types, this paper will focus only on cocaine, crack, heroin, marijuana, and meth offenses. Offenders still face a lower and higher threshold amount for each drug with a 1:10 weight ratio between thresholds. In this paper, all analysis considers only the higher threshold, which has stronger bunching, more severe punishment increases for eligible cases, and for which 70% of all federal trafficking cases are eligible.<sup>5</sup>

Cases with drug quantities at or above the MM threshold weight may not necessarily be charged with a mandatory minimum. But being charged at the weight opens the possibility for the prosecutor to impose a MM charge, meaning hitting the threshold weight significantly increases prosecutor bargaining power. Some cases have initial charges that apply the MM filing, but through plea negotiations, do not carry a mandatory minimum in the final charges. In my data, I only observe

---

<sup>2</sup>It's worth noting that in a brief appendix analysis, Tuttle (2023) finds that crack does indeed exhibit some sentence length effects if other years are included.

<sup>3</sup>Note that prosecutors cannot sum weights across drug types to get a MM charge.

<sup>4</sup>While I cannot directly identify previous offense types, I can control for "serious" prior offenses in general. Fatal or serious injury crimes can be directly identified and excluded from the analysis.

<sup>5</sup>Many trafficking cases have a charging weight close to zero. Thus, the lower MM weight threshold has a far less prominent bunching point.

final charges.<sup>6</sup> For this reason, I do not focus on the MM charges themselves in my analysis but only consider drug quantity and the sentence length. This ensures I do not leave out cases that do not show MM charges in the final charge data, but were still substantially impacted by MM eligibility during the plea bargaining process.

Furthermore, MM sentences are often non-binding, with many convictions receiving sentences above or below the minimum specified sentence length. Sentences below the mandatory minimum can occur if a defendant provides “substantial assistance” in the prosecution or investigation of another offender,<sup>7</sup> if the offender is eligible for the safety valve provision, or if a lower sentence is negotiated through plea bargain. Defense can apply for the safety valve provision if the defendant in question has a sufficiently low criminal history record.<sup>8</sup> I directly test for MM effects for cases filed under the safety valve provision.

Because MM eligibility is determined at precise weight amounts, it creates incentives for legal actors to manipulate drug weights if they are able. Evidence of this is found in significant bunching of cases right above the threshold weight, illustrated in Figure A.1. How does manipulation occur? Prosecutors connect defendants to other offenders or larger organizations, increasing the available evidence of drugs to charge a person with. Tuttle (2023) and Cooper (2023) each give a more detailed explanation of this process, and assume that manipulation only occurs upwards, meaning that case weights are “rounded up” so to speak. This assumption is supported by the presence of the manipulation effect that I find in this paper.

## **United States Attorneys**

Because I use US Attorney spells as a source of variation in bunching, I provide a brief background for their role in the charging process. United States Attorneys serve as the chief federal law enforcement agent within their district. There are 93 US Attorneys serving at all times, one for each district. US Attorneys are typically appointed by the president of the United States and serve until they choose to step down or are asked to resign. Resignation requests often occur after a new president is sworn in, but may also occur within presidential administrations. In times of vacancy, an assistant attorney already serving in that district fills the leadership role and is considered the Acting US Attorney. For extended vacancies, the US Attorney General may also appoint an interim attorney to fill the leadership role until a new presidential appointment.

---

<sup>6</sup>Rehavi and Starr(2014) observe initial and final charges for their data, but it does not include drug cases. To my knowledge, data containing initial and final charges for federal drug crimes has not been made available to researchers.

<sup>7</sup>See USSG § 5K1.1

<sup>8</sup>Before the 2018 First Step Act, this was only defendants with one or zero criminal history points. The FSA expanded this provision to include those with four or less points and only counted past crimes that were two or three point offenses. See USSC (2019) for details.

Attorneys are given immense discretion to dictate the focus and procedure of prosecution within their district. In the Principles of Prosecution section of the United States Justice Manual, it states that “...individual United States Attorneys are required to establish their own priorities (in consultation with law enforcement authorities), within the national priorities, in order to concentrate their resources on problems of particular local or regional significance.”<sup>9</sup> Other sections of the Justice Manual describe the US Attorney as having “the broadest discretion in the exercise of such authority” in relation to prosecuting criminal matters.<sup>10</sup> This implies US Attorneys have significant impacts on the types of cases that are prosecuted and the manner in which prosecution should occur. This appears to hold true in practice; in my discussion with a number of federal prosecutors, they described the US Attorney in office as having a significant impact in the day to day operations of the Assistant US Attorneys.<sup>11</sup>

This broad discretion in prosecution leads to variation in bunching across US Attorney spell. In my data, I find bunching variation both across and within districts. As shown in Table A.2, the US Attorney bunching propensity measure has a high standard deviation of 12.84. I also calculate the standard deviation within districts to ensure variation is not only driven by geographic differences. The within district standard deviation is 7.45, a number still high relative to the overall mean. Likewise, the table demonstrates variance in within district bunching by reporting mean maximum and minimum bunching scores per district, with the maximum mean at 16.41 cases and the minimum at 2.752 cases. I also show that variation is not driven only by temporal differences; Figure A.3 displays bunching for each district averaged across each US Attorney in the district. The figure shows high variation in bunching across districts.

### III Data

The primary drug case data is provided by the United States Sentencing Commission (USSC) and includes all federal drug trafficking cases from 2010 to 2021. Data is at the case-individual level. I restrict the data to the five most prevalent substances subject to MM sentencing: powder cocaine, crack, heroin, methamphetamine, and marijuana.<sup>12</sup> I also restrict the data to the primary racial

---

<sup>9</sup>See section 9-27.230.

<sup>10</sup>See section 9-2.001. This section also gives specifics about for which aspects of prosecution US Attorneys are allowed to exercise their discretion. This includes authorizing prosecution and determining the manner of prosecuting and deciding trial related questions.

<sup>11</sup>I spoke with a handful of US Attorneys and Assistant US Attorneys about the impacts of who is in the role of US Attorney, who are kept anonymous by request. While responses varied, the general consensus is that US Attorneys have a lot of flexibility in what types of cases should be prosecuted and how prosecution should be carried out. Conversations with legal scholars have told a similar story.

<sup>12</sup>The other three substances subject to MM law are PCP, LSD, and fentanyl. These had too few observations for any meaningful analysis.

groups of study: White, Black, and Hispanic. For the intensive margin analysis, the data is further restricted to include only cases with primary drug weights at 20% to 180% of the threshold weight. This gives enough data to fit distributions on, but excludes the 5-year threshold and extremely high weight cases that are likely less similar to cases near the 10-year threshold. This gives a total of 44,626 observations.

The USSC data provides a rich set of defendant and litigation details. Information on the defendant’s sex, race, age, education, and citizenship is included. The data also contains specific statutes that are charged, the drug weight given in grams or an approximation of the weight, any factors that increased or decreased the sentence length above or below the prescribed guideline amount, whether the case was plead or not, and sentence length the defendant received.

Drug weight measurement is complicated by two factors: multiple drug types and ranges of weights. Around 27% of cases involve a defendant charged with multiple drug types. In regards to the weight threshold for a MM charge, weights are to be considered separately. Thus, I consider the primary drug type for these cases, which is the drug with the highest weight relative to the 10-year threshold. However, prosecutors may sum multiple drug type weights together when determining the base offense level, implying more drug types are likely to increase sentence length. To control for this, I include the number of charged drugs as a control in some of the main specifications. I also provide discontinuity results after controlling for the weights of other drug types to ensure results are not driven by differences in multiple drug carrying behavior. This is done for the full sample and each racial group and is reported in the robustness section. For about 20% of cases, defendants are charged with a range of weights rather than one precise measure. These cases are omitted from the main analysis, but are considered across three different measurement schemes in the robustness checks.

Table A.1 gives summary statistics for several key variables in the drug data. These statistics are presented for the full sample as well as split for each racial group. The table shows that each of the five drug types have a sizable number of cases, with heroin cases making up the fewest percent of cases at 14.3% and cocaine making up the most with 26.1%. Drug type is highly correlated with race, suggesting primary specifications for race heterogeneity should control for the primary charging drug. The table also shows that Black and White defendants have very similar criminal history points, while Hispanic defendants have much lower criminal history on average. I also consider conspiracy charges as these are a primary mechanism by which drug weight manipulation occurs (Lynch 2016, Tuttle 2023, Cooper 2023). Across all three racial groups, the proportion of cases with a conspiracy charge is nearly identical. This helps alleviate concerns of unequal opportunity for manipulation across races.



In the extensive margin analysis, I use a unique, hand collected data set of US Attorneys from years 2013 to 2020. The data set includes all US Attorneys from each district, including presidential appointed attorneys, Attorney General appointed attorneys, and acting and interim attorneys that took the leadership role between appointments. This data is used to construct a measure of bunching propensity by US Attorney administration, which is key for identifying racial disparities in drug weight manipulation. This data is gathered from a number of sources including direct correspondence from US attorney district offices, US attorney district office websites, Wikipedia, and news articles. The data set includes the US attorney’s name, nomination date (if applicable), confirmation date, and date out of office. Dates are all recorded at the monthly level to match the drug data, with an attorney being counted as acting that month if the days served are greater than or equal to 16.<sup>13</sup> I match this data with the above USSC drug data to analyze manipulation effects for cases between June 2013 and September 2020. I also further reduce the sample to only include cases with weights between 50 percent and 150 percent of the threshold weight.<sup>14</sup> This gives a sample of 12,731 cases.

Table A.2 gives a few key statistics about US Attorney administrations. For the study period of 2013 to 2020, each district had an average 3.168 attorneys serve in the position with a total of 282 different attorneys. Each attorney served an average of 48.28 months and prosecuted 134.2 drug trafficking cases with weights between 50 percent and 150 percent of the threshold weight. The mean bunching propensity measure is 9.75 but with a standard error of 12.84, indicating high variance between administrations. To further illustrate the variance in bunching, I consider the maximum and minimum bunching propensity measures within each district. The mean maximum propensity is 16.41 and the mean minimum propensity is 2.752.

## IV Empirical Strategy

### Intensive Margin

To accurately estimate the causal effects of MM sentencing on the intensive margin, I need to know what the discontinuity looks like without any manipulation. Thus, my strategy is to create counterfactual distributions of sentence lengths over charged drug weight by extrapolating the distribution of unmanipulated regions into the regions near the eligibility cutoff where manipulation is present. I fit two separate distributions on either side of the cutoff. The fitted distributions are determined only by cases that are assumed to be unmanipulated. I then look at the extrapolated

---

<sup>13</sup>For some attorneys, dates of entry and exit are only available at the month level. In these rare cases, I default to the incoming attorney being the acting attorney on the month of overlap

<sup>14</sup>I do this to reduce noise in my estimation of US Attorney bunching effects. Because I am not fitting trends in this analysis removing cases unlikely to be manipulated increases estimation precision.

points right at the threshold weight. The distance between these two points gives the legal effect - the causal effect of MM law in absence of manipulation activity.

I then compare these fitted distributions against the actual sentence lengths. Specifically, I compare the point extrapolated from the right-hand side at the threshold weight against the actual sentence length at the threshold weight. Essentially, I am comparing the counterfactual sentence length at the threshold weight versus the observed sentence length at this point. The distance between these two points gives the manipulation effect. If the manipulation effect survives the inclusion of controls, it suggests that case manipulation leads to higher sentence lengths beyond the increase that comes from the legal effect.

I also look for a discontinuous drop in sentence length for cases before the cutoff. If prosecutors manipulate cases that are already likely to have higher sentences, the “worst cases,” then the average sentence where they are manipulated from will be lower than the rest of the trend. Note that this is analogous to a missing mass argument. This missing sentence amount is not formally estimated as a comparison against the manipulation effect since the number of cases varies drastically at the bunching point compared to at other charging weights.<sup>15</sup> However, I note its presence as it is evidence that prosecutors are indeed manipulating cases upward in order to increase sentence lengths for certain defendants.

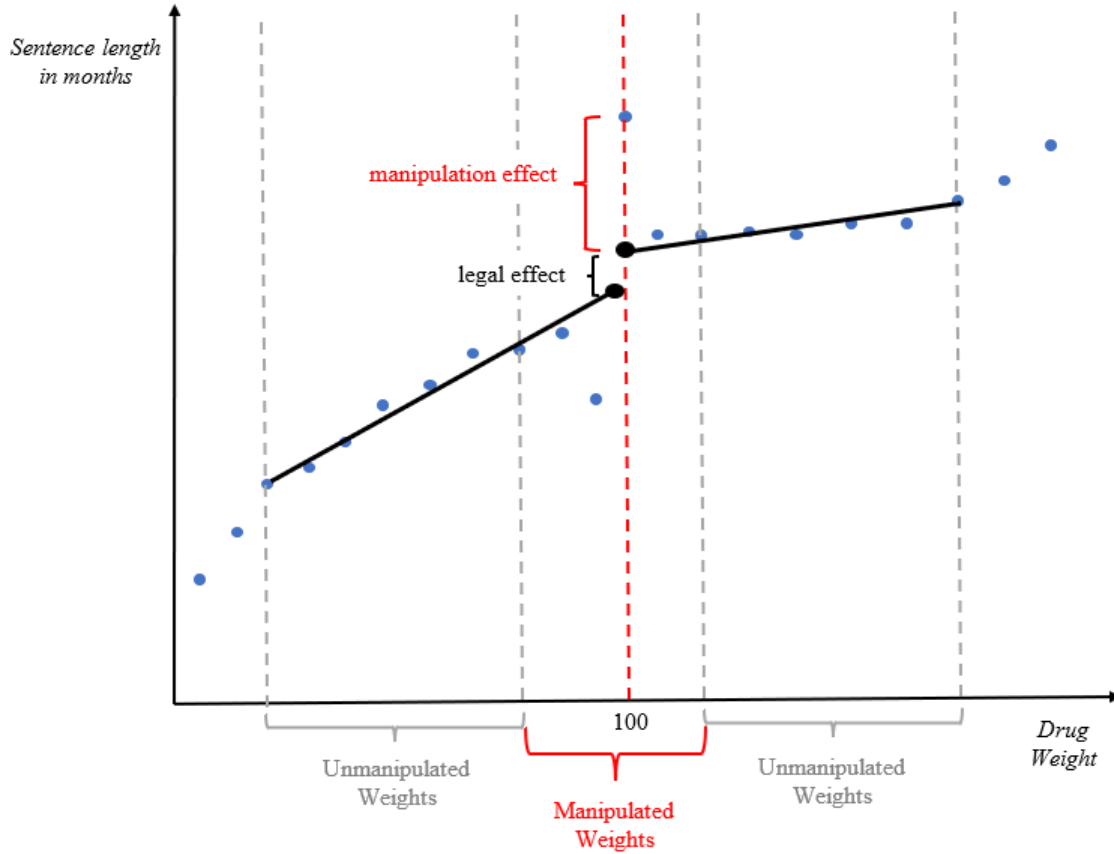
Figure 1 illustrates the idea of this strategy. The blue points represent the actual sentence length distribution for the group being considered. Dashed vertical lines show the manipulated and unmanipulated regions. The black lines give the fitted distributions which are extrapolated into manipulation regions. The large black points give the discontinuity absent manipulation. The distance between the top point and actual sentence length at the threshold gives the manipulation effect. I do not label the missing sentence amount since it is not estimated, but it can be observed as the distance between the fitted sentence just to the left of the threshold and the actual sentence lengths near this weight.

To implement this empirical strategy, I must first determine which regions are manipulated and which are not. There are two primary ways in which manipulation regions have been detected and determined in past literature. The first is a formal test developed in Frandsen (2017) and practiced in Goncalves and Mello (2021) used to detect changes in the distribution attributed to manipulation. This method is unlikely to work well in this setting because small scale manipulation likely occurs as prosecutors or law enforcement round to whole numbers. The second method is simply using visual inspection, as done in Saez (2010), Chetty et al. (2011), and Kleven and Waseem

---

<sup>15</sup>Ideally, I would multiply the missing sentence amounts by the missing mass in cases and compare this value against the product of the manipulation effect and the excess mass at the threshold weight. I refrain from doing this since missing masses in cases left of the threshold weight cannot be precisely estimated.

**Figure 1:** Empirical strategy illustration



*Notes:* This figure depicts the empirical strategy used to estimate the MM impacts and decompose them into legal and manipulation effects. The black lines represent the fit over the unmanipulated weights and extrapolated into the manipulated weights. The discontinuity between these fits at the threshold is the legal effect. The gap between the predicted sentence from the right-hand side of the distribution and the actual sentence length is the manipulation effect.

(2013). This approach works when there is a clear and obvious missing mass in the distribution that is supplying the observations at the bunching point. Where the missing mass begins is assumed to be the beginning of the manipulation region. To tease out the missing mass area, I fit a fifth order polynomial with fixed effects for each 10 percentage point round figure over the main analysis weights; 20% to 180% of the threshold weight. Figure A.2 shows this fit distribution in comparison to the actual case density. Missing mass appears following the 70% value and continues up until cases just before the MM threshold at 100%.

One common way to check whether the manipulation region is correctly specified is to compare the excess and missing mass amounts, which should be equal. The excess mass at the bunching point is just under 2,385 cases. The missing mass from 70% to 99% is about 1,129. The smaller missing mass may be generated by two sources. First, when looking at each drug individually, meth

has a manipulation range that is clearly wider than the other drug types (closer to 60%). Second, the missing mass estimation does not include cases at 70%, 80%, or 90%. These round numbers are omitted as rounding bunching occurs at these points as well. Significant case reduction may occur from cases at the 80 percentage point cases.

Based on the comparison between the fit and actual case distribution, I consider cases between 70% and 105% to be within manipulation ranges for my main analysis. This means the left-side regression is fit using cases between 20% and 69% of the threshold weight, while the right-side regression is fit using cases between 105% and 180%. Each of these fit predictions are then extrapolated in to the manipulation region to create counterfactual distributions, or the trend of sentence length absent manipulation. In my robustness checks I consider the results under many specifications with different cutoffs for the left-side manipulation region. There is a tradeoff in setting the manipulation region cutoff; cutoffs further from the threshold are less likely to be biased since they are less likely to accidentally include cases that actually are manipulated, while windows closer to the threshold have more data to fit on and are thus likely to be more precise. For this reason, it is important to show that results are similar across many different set cutoffs. This is analogous to testing a regression discontinuity with different bandwidth sizes.

This strategy can be thought of as a regression discontinuity design with discrete running variables, where the discontinuity is estimated from predicted values. Following Gelman and Imbens (2019), my analysis assumes a functional form that is linear, though I include some quadratic fit predictions in my robustness checks. I use standard errors of the prediction to create confidence intervals for each fitted value to assess inference. The legal effects are considered statistically significant if the two confidence intervals do not overlap. The manipulation effect is considered statistically significant if the actual sentence length falls outside of the confidence interval from the right-side fitted regression.

The key identifying assumption of this empirical design is that without manipulation, the distribution of sentence lengths for cases in manipulation regions would have followed the counterfactual distributions fit using unmanipulated cases. Note that this assumption is standard in a traditional bunching design, only there it is typically involving the variable being bunched, which in this context would be cases. Now I assume that the relationship between sentence length and charging weight can be predicted using the polynomial coefficients among unmanipulated cases.

A second important assumption is that prosecutors are not manipulating cases from low weight amounts (around 20%-50% of the threshold weight) up to the bunching point. This assumption is supported in three ways. One, drug weight manipulation primarily occurs from additional fact finding on the part of the prosecutor (Lynch, 2016; Tuttle, 2023). In Cooper (2023) and Online Ap-

pendix D of Tuttle (2022), costs to evidence gathering are assumed to increase as the manipulation quantity increases. Thus, if the distribution of available evidence used to manipulate weights is relatively uniform across seized drug weights, the probability of manipulation should be highest for cases just below the threshold weight. This manipulation probability will decrease for lower seized weights, indicating manipulation is more likely to occur for cases near the threshold. Two, Cooper (2023) assesses decreases bunching caused by a policy and finds increased mass appear primarily for cases near the bunching point. This indicates these marginal cases were mostly manipulated near the threshold weight.<sup>16</sup> Three, missing sentence lengths appear at weights just below the threshold.

## Extensive Margin

To provide evidence of race disparities on the extensive margin, I utilize the variation in bunching propensity by US Attorney administration as a pseudo-random measure of the probability of being assigned bunching. The idea is that non-bunching and low-bunching administrations serve as reliable counterfactual distributions compared against high-bunching administrations.<sup>17</sup> This is a similar strategy as described in Frandsen (2017) and practiced in Goncalves and Mello (2021). In their analysis, Goncalves and Mello (2021) estimate a binary measure of traffic citation bunching and compare bunching versus non-bunching police officers. Rather than trying to determine a cutoff between bunching and non-bunching administrations, I use a residualized, continuous measure of bunching propensity to identify prosecutor-driven racial disparities. This approach has a flavor of judge-leniency instrumental variables as seen in a large number of law and economics papers (Kling 2006; Aizer and Doyle 2015; Mueller-Smith 2015; Bhuller et al. 2020; Di Tella and Schargrodsky 2013; Dobbie et al. 2018). However, I don't use the propensity measure as an instrument but consider it a proxy for the defendant's probability of being charged at the bunching weight.

To estimate racial disparities, I employ the following model:

$$\begin{aligned}
Y_{idmt} = & \alpha + \beta_1 Black_i + \beta_2 Hispanic_i + \beta_3 bunch\_score_{idmt} \\
& + \beta_4 Black_i \times bunch\_score_{idmt} + \beta_5 Hispanic_i \times bunch\_score_{idmt} \\
& + X_i\gamma + \lambda_d + \kappa_t + \eta_m + \epsilon_{idmt} \quad (1)
\end{aligned}$$

with  $\beta_4$  and  $\beta_5$  being the primary coefficients of interest. These give the differential effects of

<sup>16</sup>Following Cooper (2023) I graph the weight distributions before and after the policy for the primary affected group, but on a wider distribution to check which cases are being manipulated. I find excess mass appears mostly near the threshold, with some excess mass appearing between 50% and 60% of the threshold. Figure available upon request.

<sup>17</sup>Note that for this analysis, I only consider cases from 50% to 150% of the threshold weight as I am only concerned with the effects of manipulation and want to minimize fluctuations in other parts of the distribution.

increasing bunching propensity on Black and Hispanic defendants compared to White counterparts.  $X_i$  gives observable characteristics of the case and of the defendant. These include the drug type, the total number of criminal history points applied in the case, the sex of the defendant, defendant age and age squared, a binary for whether the defendant completed some or graduated from college, and a binary for whether the defendant is an illegal alien. It also contains a measure of the percent of cases with a White defendant per US Attorney spell to control for differences in case composition and prosecution opportunity across race. I likewise include year, month-of-year, and district fixed effects. The fixed effects control for temporal shocks, seasonality, and district differences in bunching behavior.

To construct the bunching propensity measure, I use a leave-out, residualized mean of US Attorney bunching decisions, similar to the approach used in Dahl et al. (2014), Dobbie et al. (2018), and Arnold et al. (2018). I regress the bunching decision on year, month of year, and district fixed effects. I then take a leave-one-out mean of the residuals across each US Attorney, where each individual’s bunch score value is a mean across all other cases for the US Attorney in office during the individual’s case timing and location. Controlling for time and location fixed effects is necessary since US Attorney assignment is not truly random. High or low bunching attorneys may be assigned to a district at a specific time in response to drug crime behavior in the area.

The primary assumption of this method is conditional random assignment of bunching propensity to a defendant. Another way of saying this is that whether a defendant is prosecuted by a high bunching or low bunching administration is effectively random after controlling for area and time fixed effects. This implies individuals are not changing their criminal behavior based on the current or recent US Attorney prosecution patterns. This seems reasonable; to have a meaningful impact on criminal behavior, criminals would need to be aware of the US Attorney’s position on drug prosecution and that knowledge would need to have a strong enough incentive to change production or transportation activity. Beyond this, many federal cases pass through the state system first, meaning there is uncertainty for the defendant about which level they will be prosecuted at. A bigger threat comes by way of detection and law enforcement activity. US Attorneys work in close contact with members of the FBI and sometimes with the US Marshall’s Service or the DEA. A high bunching attorney may also encourage specific types of drugs be targeted or more arrests in general.

While I cannot test directly for changes in criminal or law enforcement, I can broadly test for selection on observables by regressing bunching propensity on defendant characteristics. Table A.3 gives  $F$ -Statistics and tests for regressions run on observables. The first column tests just for

bunching just as a binary variable. The joint  $F$ -test for this specification returns an  $F$ -value of 7.01 and a  $p$ -value of 0.000, indicating non-random selection of who has their case manipulated. The next three columns then regress the residualized bunching propensity measure on observables. The resulting  $F$ -statistics decrease significantly from the bunching decision test with values between 0.75 and 1.30, indicating conditional random assignment to high or low bunching US Attorney spells. The  $p$ -value also increases and is insignificant across each tested sample. To test changes in law enforcement, I consider the main specifications again but drop attorney spells with especially low or high cases per month. This is discussed in more detail below in the robustness checks section of the results.

Finally, it may be that racial disparity is driven by unobserved characteristics of the case that are correlated with race that change the prosecutor’s bargaining power. For instance, if racial minorities happened to sell to minors more often, that may increase the expected sentence length but would be unobservable in the data. Randomization by attorney propensity may help control for this, but it is possible that low bunching attorneys ignore these effects while high bunching ones exploit them. I argue the racial disparity is not driven by such factors by comparing departure rates between race groups. Departure means an individual is charged outside of the recommended sentencing guidelines for some specific reason. A departure can increase the sentence if the defendant has what are called aggravating factors; these include things like cruelty, involving a minor, or having a leadership role. Departure can decrease a sentence if the defendant has mitigating factors; these include things like having a minimal role in selling or being a minor. Thus, if unobserved case characteristics were driving the bunching disparity, the rate of departure would be different by race. I find this is not the case. Furthermore, I find that controlling for departure does not effect the statistical significance of the results nor does it substantially affect magnitudes of the estimates.

## V Results

### Intensive Margin

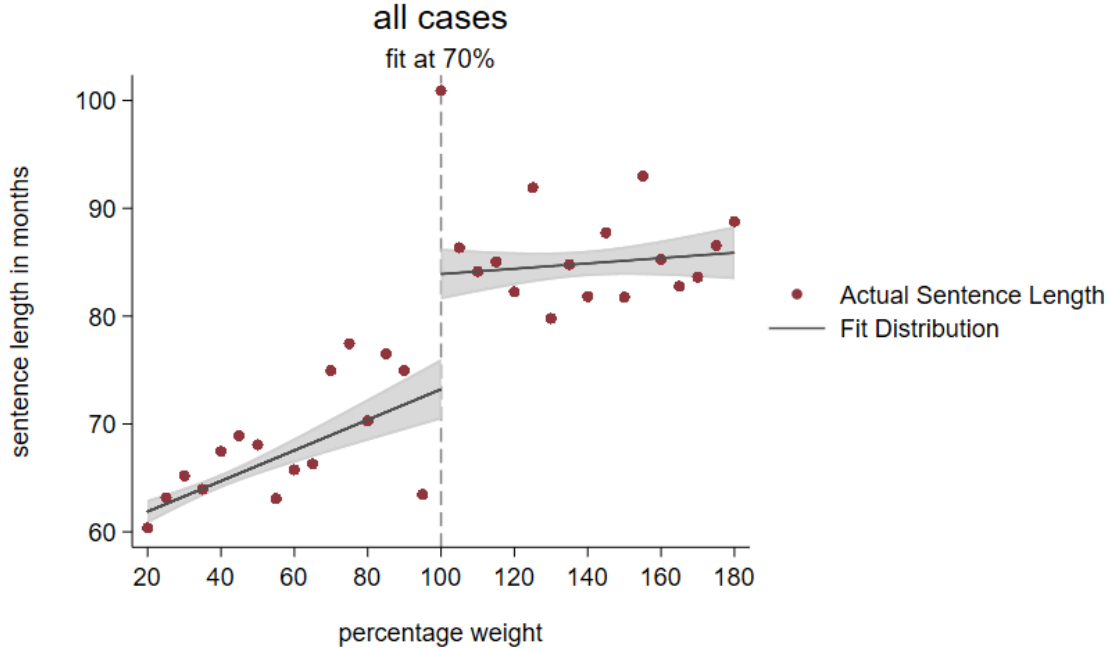
Before presenting the main race disparity results, I first discuss findings for the analysis aggregated across all cases. I then show then compare legal and manipulation effects across racial groups. After these main results, I briefly discuss heterogeneity across drug type and criminal history group. For each of these analyses, my preferred specification is a linear functional form with the manipulation region being defined as cases between 70% and 105% of the threshold weight and without controls. The simple model is preferred so that any mechanisms driving disparities between groups is not controlled for. I also provide analysis with varying sets of controls, varying

fit cutoff values, and with quadratic fits in the main results or in the robustness section.<sup>18</sup> I then include a brief section on mechanisms for manipulation effects.

### *Overall Effects*

Figure 2 shows the discontinuity aggregated across all cases. The figure highlights three important findings. First, the fit values approximate sentence lengths quite closely within the manipulation range except for cases closest to the threshold weight. Second, the legal effect of mandatory minimums is significant, with an increase of 9.63 months. This is a 15.9% increase over the mean at 99% threshold weight. Third, the figure shows a large manipulation effect with the actual sentence length at the bunching point far higher than the fit value. I estimate a difference of 16.08 months between these points. And fourth, the missing sentence appears to be local to cases just left of the threshold weight. The presence of missing sentence lengths gives evidence that prosecutors are manipulating cases up to the threshold weight. The size of the missing sentence is estimated to be 11.27 months. Each of these three effects are statistically significant.

**Figure 2:** Discontinuity analysis for all cases



*Notes:* This figure gives the main discontinuity analysis for all cases with linear fits through points before 70% of the threshold weight and extrapolated past 70% to the threshold from either side. Sentence lengths are binned by 5% except at the threshold weight, which only reports average sentences for cases charged with exactly the threshold weight.

<sup>18</sup>Many sub group results, particularly drug type-race intersections, are not presented here for space. Results for specific subgroups are available upon request to the author



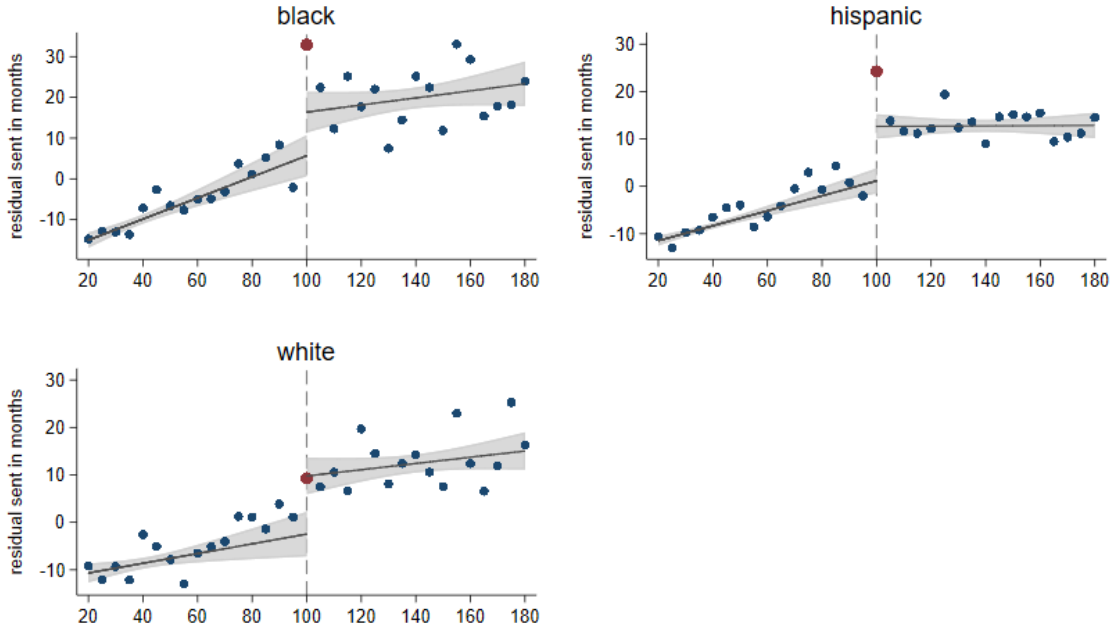
Panel A of Table A.4 reports these results along with results for two different control schemes. The baseline controls include criminal history points, drug type, and racial group. The additional controls include the defendant’s age, their education, whether they are female, a binary for gun use, and whether the case went to trial. The fixed effects are at the district and year level. The size of the manipulation effect decreases under the control schemes to 11.11 and 7.875. This means that even with the full set of controls, there is still a sizable sentence premium that cannot be accounted for by the included observables. If this estimate were to be interpreted as the manipulation effect free from selection, it would imply that cases that are manipulated to the 10-year mandatory minimum threshold receive a causal increase to their sentence length of 16.751 months (this is the legal + manipulation effects after controls). Statistical significance persists through each specification.

### *Racial Disparities*

Figure 3 illustrates the discontinuity analysis separated by race with sentence lengths residualized on the baseline controls, which in this case are criminal history points and drug type. These regression results are also reported in Table 1. There are two key results presented here. First, the legal effect is nearly identical across each race group, with Black defendants having the smallest legal effect at 10.28 months, and White defendants with the largest at 12.12 months. It’s worth noting that by percentage of mean sentence left of the cutoff, the effects vary more, with Hispanic and White defendants experiencing a significantly larger impact from the legal effect compared to Black defendants. However, the results still imply that, absent any manipulation, the causal impact of mandatory minimum eligibility is effectively equal across race in terms of level sentence lengths.

The second key finding is that the manipulation effects are driven exclusively by racial minority defendant cases. Black defendants at the weight threshold experience sentence lengths 14.81 months higher than cases just to the right of the threshold, even after controlling for criminal history and drug type. Similarly, Hispanic cases have a manipulation effect of 10.85 months in this specification. But White cases have a manipulation effect of essentially zero. These effects shrink under the full control scheme, with manipulation effects of 7.15 months and 5.11 months for Black and Hispanic defendants, respectively. This provides evidence that prosecutors choose to manipulate cases more frequently for minority cases than White ones, corroborating findings from Tuttle (2023) and Cooper (2023). These results also indicate that minority defendants who are manipulated receive a sentence premium, meaning the causal impact of mandatory minimum eligibility is higher for minority cases than White ones in the presence of manipulation.

**Figure 3:** Race comparison - discontinuity using residualized sentence lengths



*Notes:* This figure gives the residualized sentence discontinuity analysis by racial group with linear fits through points before 70% of the threshold weight and extrapolated past 70% to the threshold from either side. Sentences are residualized on criminal history points and drug type. Residual sentence lengths are binned by 5% except at the threshold weight, which only reports average sentences for cases charged with exactly the threshold weight.

The results can also be considered at each drug-race intersection. However, these samples become smaller and noisier, leading to less precise estimates. The overall story is the same though - legal effects are similar across groups, though estimates tend to be strongest for whichever race has the most cases for a drug type. In extreme cases, one or two race groups make up almost all of a drug type's cases, such as Black defendants for crack cases. In these cases, legal effects are only present for the high case group. Manipulation effects are driven primarily by racial minorities in cocaine and heroin cases, though some smaller effects exist for other minority race-drug combination. These results still highlight that the law effects defendants similarly, but manipulation decisions only penalize Black and Hispanic defendants.

#### *Drug Type and Criminal History*

Here I briefly consider heterogeneity in effects across drug type and criminal history. While these factors are controlled for in the residualized results, I consider here whether drug types are a key mechanism for manipulation decisions. Due to its importance in drug sentencing history and interaction with Black defendants, I especially consider whether results are driven primarily by crack cases. I then repeat the analysis for low level offenders.

**Table 1:** MM effects - race comparison

	(1)	(2)	(3)
	black	hispanic	white
legal effect	10.63	11.43	12.22
left side 95% CI	[0.494, 10.52]	[-1.552, 3.683]	[-7.196, 2.118]
right side 95% CI	[11.53, 21.38]	[10.11, 15.20]	[6.047, 13.62]
manip. effect	16.56	11.61	-0.462
fitted mean at 99%	7.54	3.16	-1.71
fit value	70%	70%	70%
N of obs fit on	11205	15296	7190

*Notes:* The dependant variable is sentence length residualized against criminal history points and drug type. All specifications are discontinuities based on local linear fits. 95% confidence intervals are presented for each fit regression on either side of the cutoff. These are calculated using the standard error of the predicted expected value, and significance is determined as no overlaps between these two intervals. The legal effect is the regression discontinuity between the two extrapolated fits at the 10-year MM cutoff. The manipulation effect is the difference between the right-hand regression fit and the actual sentence length at the threshold weight. Fit value represents the cutoff for where extrapolation begins. In this case, all specifications have left-hand regressions fit on cases with weights between 20% and 70% of the threshold weight, with extrapolation occurring from 71% up to the cutoff.

Figure A.5 gives the main discontinuity graphs across each drug type. The legal effect is estimated to be between 9 months and 14 months for each type besides for meth, which has a large legal effect of 20 months. The manipulation effects are driven almost entirely by only two drug types: cocaine and heroin. It's noteworthy that crack cases do not exhibit bunching, a result that corroborates findings from Bjerk (2017b). This may come from the fact that overall sentence lengths are much higher for crack cases than any other drug type. So while mandatory minimums seem to have less impact on crack compared to other drugs, crack cases are still punished at a higher level it seems.<sup>19</sup> For the preferred specification, the manipulation effect for cocaine and heroin cases are 26.15 and 27.31 months, respectively. However, when controlling for race and other defendant and crime characteristics, the manipulation effect shrinks to just over 5 months.

To measure low-level offender effects, I consider effects for two subgroups: defendants who have had no criminal history contact of any sort, and those who have had contact but have 0 or 1 criminal point, which relates to crimes with sentences less than 60 days. Because there are policies designed to protect first-time and low-level offenders from MM sentences, the expected effect of MM eligibility may be low for these two low history groups. Bjerk (2017b) reports this expected result of almost no effects for low level defendants.

I find that individuals with low criminal histories are still subject to both legal and manipulation

<sup>19</sup>This may stem from differences in criminal behavior, systemic forces, or discrimination. Explaining the sentence length disparities across drug types outside of MM eligibility is beyond the scope of this paper.

effects. Figure A.4 shows the regression discontinuity graphs for these two criminal history groups. Individuals with no prior criminal justice interactions experience a legal effect increase of 6.721 months and a manipulation effect of 11.49 months. Defendants with 0 or 1 points experience a legal effect of 10.18 months and a manipulation effect of 10.03 months. Both of these groups exhibit a relatively small missing sentence amount, suggesting that selection may not be a primary driver of the manipulation effect. However, when I subset for cases that receive the safety valve provision, this manipulation effect disappears. Taken together, the results suggest that legal protections do not fully shield low history offenders from higher sentences, nor do they protect them from effects driven by case manipulation in non-safety valve cases.

### *Intensive Margin Mechanisms*

If prosecutors are selecting cases to be manipulated upward, the cases most likely to be manipulated are those that either have low cost to manipulation, or high benefit for achieving the 10-year MM threshold (likely through higher sentence lengths).<sup>20</sup> Both of these channels are plausible. Cooper (2023) provides evidence that prosecutors tend to bunch more when costs to bunching are reduced. Here, I give evidence that prosecutors manipulate defendants with higher sentence length potential.

First, the presence of missing sentence lengths provides evidence that prosecutors manipulate cases with relatively high average sentences. These significant deviations in sentence trends only appear for cases in which the manipulation effect is present and large. Specifically, the regression analysis for heroin cases and for Black and Hispanic cases, given respectively in Figures A.5 and 3, show significantly lower sentence lengths for cases between 95% and 99% of the threshold weight.

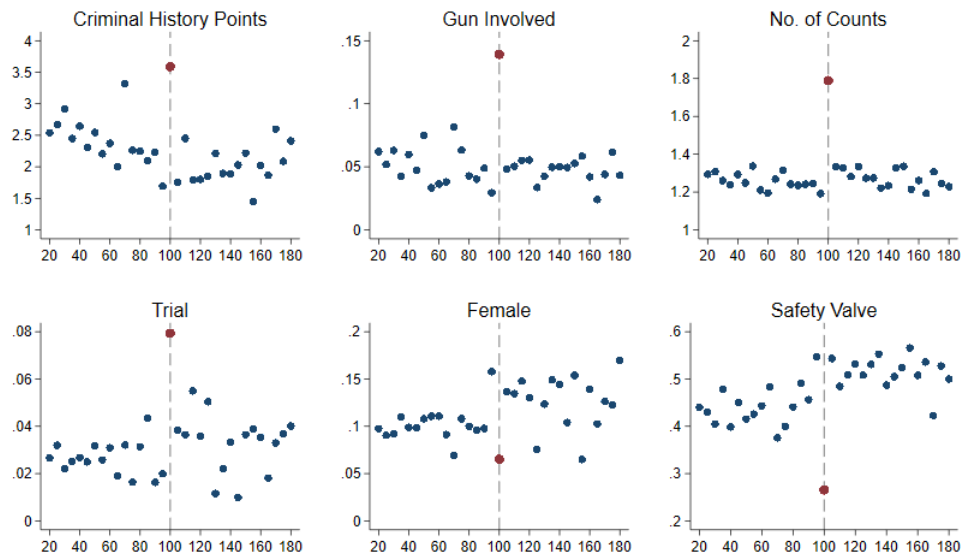
Second, if prosecutors are systematically choosing defendants to get higher sentence lengths through manipulation, cases at the threshold should have similar bunching type patterns across other observable factors that are highly correlated with sentence length. One way to think of this is that prosecutors might manipulate weights for the defendants with the “worst” characteristics, or those that are most likely to increase a defendant’s sentence length. Thus, I consider how charging weight correlates with six factors strongly associated with higher sentences: criminal history points, whether a gun was involved in the crime, the number of counts a defendant is charged with, whether the case went to trial, whether the defendant is female, and whether the safety valve provision was applied at sentencing. I graph these factors against charging weight for the cases with the highest bunching effects - cocaine, crack, and marijuana cases for Black and Hispanic defendants.

---

<sup>20</sup>Note that this assumes that prosecutors generally prefer higher sentence lengths to lower ones. This may not be true in every case, but it is not inconsistent with proposed models, whether prosecutors seek to minimize errors, maximize convictions, maximize sentence lengths, or some combination of these. See Landes (1971), Grossman and Katz (1983), and Silveira (2017).

Figure 4 presents the results. The figure shows significant selection effects across each of these factors. The mean at the MM threshold is a significant outlier compared to other charging weights: for criminal history the bunch mean is 54.9% higher than the non-bunch mean, 158.9% higher for gun involvement, 41.3% higher for number of counts, 167.8% higher for trial, 40.3% lower for female, and 41.9% lower for safety valve. Each of these variables are positively correlated with sentence length besides female and safety valve, which are negatively correlated. The results imply that manipulation occurs upward and for high-sentence cases.

**Figure 4:** Factors being selected on



*Notes:* Each graph displays a different key variable measured across drug weight as a percentage of the threshold weight. These are shown in 5% bins, except at the threshold weight which only contains cases charged exactly at the threshold weight. Each variable has a strong, positive correlation with sentence length except for whether the defendant is female or receives the safety valve provision, which correlate negatively. The point at the threshold is a different color and size to highlight its difference compared to other charging weights.

## Extensive Margin

I now consider whether racial disparities exist in MM eligibility on the extensive margin. Again, I do not consider whether minorities are charged with mandatory minimums more than White counterparts, for which there is evidence in prior literature (Rehavi and Starr 2014). Likewise, Table A.1 shows that for cases within 20% and 180% of the threshold weight, White individuals have the highest average weight. This indicates that on average, minorities are not eligible more often than White defendants, implying race gaps are likely not driven by systemic forces. Instead, I aim to measure whether prosecutors choose to manipulate cases more frequently for minority cases than for White ones. This is done by comparing bunching across races in high versus low bunching

US Attorney spells.

I first measure racial differences in bunching propensity itself using the model specified in Equation 1 with a binary measure for whether a case is at the bunching point or not as the dependent variable. Table 2 displays these results. Across these specifications, I estimate the effects for the full sample, the sample excluding attorneys with less than 25 total cases, and the sample excluding attorneys with less than 50 total cases.<sup>21</sup> I reduce the sample this way to ensure outliers are not driving results. For the propensity effects, the first four columns vary defendant controls and fixed effect combinations.

**Table 2:** Extensive Margin - Racial Disparity in Bunching

	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.00816 (0.00888)	-0.0113 (0.0116)	-0.00208 (0.0105)	-0.0167 (0.0123)	-0.0172 (0.0133)	-0.0222* (0.0127)
Hispanic	0.00806 (0.00707)	-0.0113 (0.00979)	0.0113 (0.0106)	-0.0158 (0.00971)	-0.0144 (0.0108)	-0.0165 (0.0116)
bunch_score	0.00896*** (0.000898)	0.00855*** (0.000944)	0.00783 (0.0135)	-0.00513 (0.00707)	0.00391 (0.0107)	0.00767 (0.0134)
Black*bunch_score	0.00230* (0.00123)	0.00277** (0.00124)	0.00375** (0.00141)	0.00293** (0.00141)	0.00294** (0.00141)	0.00398*** (0.00141)
Hispanic*bunch_score	0.000769 (0.000948)	0.000481 (0.00101)	0.000844 (0.00134)	0.000809 (0.00103)	0.000433 (0.00108)	0.000865 (0.00133)
Covariates	no	yes	no	yes	yes	yes
Fixed Effects	no	no	yes	yes	yes	yes
Attorney # of Cases	All	All	All	All	≥ 25	≥ 50
R Squared	0.163	0.184	0.169	0.186	0.192	0.186
N	12210	10927	8553	12210	10927	8553

*Notes:* The dependant variable is a binary variable for whether the charging weight is at the threshold weight (bunching point). The first four specifications vary whether fixed effects or covariates are included in the regression. Specifications 5-6 reduce the sample to US Attorney spells that had 25 or greater or 50 or greater drug cases within the target 50%-150% weight range. Standard errors are clustered at the district level.

Significance levels: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

The estimates indicate Black defendants face significantly higher bunching odds when assigned to a high bunching administration. Specifically, a 10 percentage point increase in bunching propensity of the US Attorney administration increases the probability of a Black defendant getting bunched by 2.79 to 3.75 percentage points over White counterparts. This represents a 28.6% to 38.5% change compared to the overall bunching mean. These estimates are statistically significant at the 5 percent level across the specifications, except the first. Consistency in magnitude and significance of estimates regardless of the inclusion of defendant characteristics indicates that ob-

<sup>21</sup>Note that the number of cases mentioned here is just the number of drug trafficking cases within the 50 percent - 150 percent weight window. Even US Attorneys with relatively short spells see many cases when including all weights and other types of crime.

servable characteristics are not driving the results. Effects for Hispanic defendants are also positive, indicating higher bunching for them over Whites, but the effects are smaller and only marginally significant dependant on specification.

Do higher bunching attorney spells drive race gaps in sentence length? To answer this, I use the same method outlined in Equation 1 with sentence length as the dependant variable. Note that this is not necessarily a causal estimate of how bunching impacts sentence.<sup>22</sup> Rather, I assess whether higher bunching attorney spells correlate with higher sentence lengths throughout the distribution, or only at the threshold weight. Table A.5 displays these results. I find higher bunching attorneys are associated with significantly higher sentence lengths for racial minorities, but only when the threshold weight cases are included. This corroborates findings in the extensive margin analysis; that race gaps in sentencing are driven by defendants who have their case manipulated to the bunching point.

## Robustness

One of the primary concerns for the main results is that using the 70% cutoff for fitting over non-manipulated cases either creates bias (if 70% is too low) or doesn't capture all the variation in the data well (if 70% is too high). To check that results and magnitudes are not driven by one specific fit, I repeat the analyses for the overall results and the heterogeneous results using a variety of different fit cutoffs. For each group, I run the main regression discontinuity again for every cutoff between weights 60% and 90% of the threshold for a total of 30 additional specifications (in addition to the 70% analysis). I present these results in four specifications - one in which all fits are aggregated and averaged into one fit, and the median, upper bound, and lower bound fits with regards to the legal effect. These are displayed for all cases together in Panel B of Table A.4. The legal effect remains largely the same, though around a month smaller for most specifications as 70% happens to be the upper-bound for the full sample. The lower bound of the legal effect is 6.362 and is still significant at the 5% level. Note that the manipulation effect doesn't change across these four specifications as the cutoff value only affects the left-side regression fit.

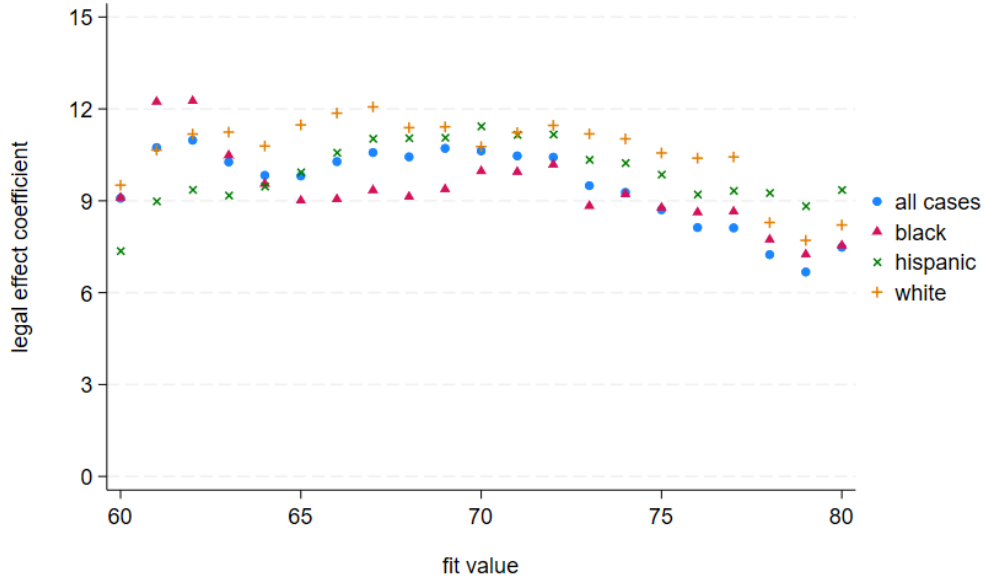
The heterogeneous effects presented in the main results are strongly consistent across different fit values. Groups with noisier data, like crack cases, tend to have higher variability of results by fit. Even still, these results tell a similar story. I present the legal effect coefficients across fit by race and for all cases in Figure 5. This graph shows the result for all 21 regressions for each group. The legal effect is consistent across these specifications, suggesting results are not driven by specific

---

<sup>22</sup>To estimate this, I would need to use US Attorney bunching propensity as an instrument for being bunched, which would not reasonably meet the exclusion restriction.

cutoff choices.

**Figure 5:** Legal effect robustness - all cases and by race



*Notes:* Each point gives the magnitude of the legal effect regression under a specific fit value. A fit value designates which area is considered manipulation region and which is considered unmanipulated. For example, at a fit value of 65, the regression is fit using all cases with weights between 20% and 65% of the threshold weight. The regression is extrapolated from 65% up to the 10-year MM cutoff, where it is compared against the right-hand regression to estimate the 65% legal effect coefficient.

I also check that the main results hold using a quadratic fit. For this check, I only use a cutoff value for the manipulation region of 80% of the threshold weight. I use a higher cutoff because at 70%, the quadratic polynomial tends to overfit the data for certain groups. Thus, to assure a reasonable fit I increase the amount of data to fit on and decrease the amount of extrapolation needed. The quadratic fits do not largely change the results, only slightly increasing or decreasing effects. Following Gelman and Imbens (2019), I refrain from using higher order polynomials, especially given concerns of overfitting across the non-manipulated region of cases.

In the main analysis, I control for the number of other drug types a defendant is charged with outside of the primary drug type. Readers may be concerned that the quantity of other drugs is driving results, especially if other drug type quantities is correlated with certain racial groups. I consider the legal and manipulation effects for the full sample and across racial groups again, now controlling for the quantity of other drug types rather than just the number.<sup>23</sup> I find effect sizes and statistical significance very close to those listed above, though with smaller manipulation effects.

Another concern may be that limiting the sample to cases with precise weights biases results if imprecise measures are strongly correlated with sentence length. I now rerun the full sample

<sup>23</sup>Table available upon request.



analysis including cases that are charged with a range of weights rather than one precise count. I use three different measures to do this - the minimum, median, and maximum of the range. These range cases are combined with the precise weight cases for a combined samples of 51,975 cases when using the maximum or median, and 56,189 when using the minimum measure.<sup>24</sup> Across all three measures, the legal and manipulation effects have similar magnitudes to main results and remain statistically significant.

To check that the extensive margin results are not driven by law enforcement changes or substantial changes in volume of cases prosecuted, I run the above analysis again omitting attorney spells with especially high or low cases per month. Note that this is different than omitting attorneys with less than 25 or 50 total cases; that restriction removes attorneys with very few total cases, mostly removing very short US Attorney spells. Case-per-month restrictions control for times in which there happens to be an especially high volume of cases entering the system, a measure that will strongly correlate with law enforcement behavior relating to drug trafficking. Table A.6 displays these results. Columns 1 and 2 omit US Attorney spell cases with the bottom 10 percent of cases per month while columns 3 and 4 omit the top 10 percent. These estimates have similar magnitudes to the main specifications and remain statistically significant.

I also consider the sentence length effects again using a binary framework of treatment rather than a continuous measure. This methodology is closer to the approach used in Goncalves and Mello (2021), where the estimates are traditional difference-in-difference results comparing non-bunching to bunching groups across race. I use two separate binary measures in this analysis. For the first measure, I distinguish bunching from non-bunching attorneys by comparing the percent of cases charged at 99 percent weight, 101 percent, and at the bunching point. I take a ratio of the bunching weight compared to all weights in this window. The US Attorney bunching treatment equals one if this ratio is strictly greater than one third.<sup>25</sup> The second binary measure simply splits US Attorney spells in half between highest and lowest bunching using the residualized continuous measure.

I also consider the sentence length effects again using a binary framework of treatment rather than a continuous measure. This methodology is closer to the approach used in Goncalves and Mello (2021), where the estimates are traditional difference-in-difference results comparing non-bunching to bunching groups across race. This results in similar findings to the main analysis above; I find high bunching spells are associated with more bunching and increases in sentence lengths for Black

---

<sup>24</sup>Table available upon request.

<sup>25</sup>This is essentially a simplification of the method described in Frandsen (2017) and used in Goncalves and Mello (2021). The main difference is I do not estimate a distribution curvature parameter  $k$  or test for the probability an attorney falls within the estimated range.

and Hispanic defendants compared to White ones, though statistical significance is marginal using this measure.<sup>26</sup>

## VI Conclusion

Mandatory minimum sentencing is a controversial practice in the criminal justice system with debate as to its true impact on criminal punishment. Concerns have largely surrounded two main points; that mandatory minimums may lead to large increases in sentence lengths for non-violent offenders and that mandatory minimum sentencing may be a cause of racial disparities in incarceration time. In this paper, I show that mandatory minimum eligibility for federal drug cases affects sentence length in two ways: through a legal effect and a manipulation effect. This decomposition is crucial for accurately estimating mandatory minimum causal effects and for understanding how policy intervention might affect disparities between cases with different drug types and defendants of different criminal history background or different racial groups.

My intensive margin analysis shows that without manipulation, MM eligibility still leads to significant increases in sentence length. The average legal effect of eligibility is a 9.626 month increase in sentence length, which is a 12.76% increase over the counterfactual mean. The level effect is fairly consistent across drug type and racial group. This means that race gaps in sentence length are not driven by some systemic aspect of the law itself. It also means that for most cases, MM eligibility increases punishment homogeneously regardless of drug type or defendant race. However, for a subset of specific types of cases that are close to the 10-year eligibility threshold weight, manipulation effects can lead to even larger increases in sentence length. When controlling for criminal history and drug type, I find that Black defendants receive an additional 7.15 months to their sentence due to manipulation. Similarly, Hispanic cases receive an additional 5.11 months from manipulation. Thus, racial disparities in sentence lengths caused by mandatory minimums are driven almost entirely by manipulation of legal actors. These effects are primarily driven by powder cocaine and heroin cases. The legal and manipulation effects are also present for low-history offenders, meaning current policy does not fully shield them from MM impacts.

To estimate effects on the extensive margin, I consider whether minorities are disproportionately pushed to the bunching point during high spells. Using a unique data set of US Attorneys, I find that US Attorney spells with higher bunching propensity cause more bunching and disproportionately higher sentence lengths for racial minorities compared to Whites. This correlation in sentence length only holds at the bunching point, meaning that high-bunching attorneys spells do not assign higher sentences to minorities in general but only do so through drug weight manipulation. This

---

<sup>26</sup>Table and measurement details available upon request.

implies that the racial disparities in manipulation effects found in the intensive margin are driven by prosecutor choices. Taken together, the results show that on both margins, the race disparities caused by MM eligibility are driven almost entirely by discretionary forces rather than systemic ones.

This paper suggests that decreasing MM sentencing impacts is likely to reduce sentence lengths for most drug trafficking cases, and would do so fairly uniformly. However, it is difficult to say how policy might impact racial disparities since race gaps are driven by prosecutor discretion. Furthermore, elimination or changes to MM sentencing may be met with compensating behavior from prosecutors or judges. It is also worth noting that this paper does not address large racial disparities in sentence lengths that are not driven from MM sentencing. Future work will be needed to understand how legal institutions and legal actors create disparities for specific defendant types.

## References

- Aizer, A. and J. J. Doyle, Jr. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *Quarterly Journal of Economics* 139(2), 759–803.
- Arnold, D., W. Dobbie, and C. S. Yang (2018). Racial bias in bail decisions. *Quarterly Journal of Economics* 133(4), 1885–1932.
- Arora, A. (2019). Too tough on crime? the impact of prosecutor politics on incarceration. Working paper, American Economic Association.
- Bhuller, M., G. B. Dahl, K. V. Loken, and M. Magne (2020). Incarceration, recidivism and employment. *Journal of Political Economy* 128(4).
- Bjerk, D. (2005). Making the crime fit the penalty: The role of prosecutor discretion under mandatory minimum sentencing. *Journal of Law and Economics* 48(2), 591–625.
- Bjerk, D. (2017a). Mandatory minimum policy reform and the sentencing of crack cocaine defendants: An analysis of the fair sentencing act. *Journal of Empirical Legal Studies* 14(2), 370–396.
- Bjerk, D. (2017b). Mandatory minimums and the sentencing of federal drug crimes. *Journal of Legal Studies* 46(1), 93–128.
- Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *Quarterly Journal of Economics* 126(4), 1593–1660.
- Cooper, S. (2023). Prosecutor tradeoffs and race: Evidence from a circuit split. Working paper, University of Connecticut.
- Dahl, G. B., A. R. Kostol, and M. Mogstad (2014). Family welfare cultures. *Quarterly Journal of Economics* 129(4), 1711–52.
- Di Tella, R. and E. Schargrodsky (2013). Criminal recidivism after prison and electronic monitoring. *Journal of Political Economy* 121(1), 28–73.
- Diamond, R. and P. Persson (2017). The long-term consequences of teacher discretion in grading of high-stakes tests. Working paper, Stanford University.
- Didwania, S. H. (2020). Mandatory minimum entrenchment and the controlled substances act. *Ohio State Journal of Criminal Law* 18.
- Dobbie, W., J. Goldin, and C. S. Yang (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review* 108(2), 201–40.
- Fischman, J. B. and M. M. Schanzenbach (2012). Racial disparities under the federal sentencing guidelines: The role of judicial discretion and mandatory minimums. *Journal of Empirical Legal Studies* 9(4), 729–764.
- Frandsen, B. R. (2017). *Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design when the Running Variable is Discrete*. Bingley, UK: Emerald Publishing Limited.

- Gelman, A. and G. Imbens (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business and Economic Statistics* 37(3), 447–456.
- Goncalves, F. and S. Mello (2021). A few bad apples? racial bias in policing. *American Economic Review* 111(5), 1406–41.
- Grossman, G. M. and M. L. Katz (1983). Plea bargaining and social welfare. *The American Economic Review* 73(4), 749–757.
- Kleven, H. and M. Waseem (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *Quarterly Journal of Economics* 128, 669–723.
- Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review* 96(3), 863–76.
- Landes, W. (1971). An economic analysis of the courts. *Journal of Law and Economics* 14(1), 61–107.
- Lynch, M. (2016). *Hard Bargains: The Coercive Power of Drug Laws in Federal Court*. New York: Russell Sage Foundation.
- Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. Working paper.
- Rehavi, M. M. and S. B. Starr (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy* 122(6), 41320–1354.
- Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy* 2(3), 180–212.
- Shaffer, H. (2023). Prosecutors, race, and the criminal pipeline. *The University of Chicago Law Review* 90(7), 1889–1965.
- Shaffer, H. and E. Harrington (2017). Brokers of bias in the criminal justice system: Do prosecutors compound or attenuate racial disparities inherited at arrest? Working paper.
- Silveira, B. S. (2017). Bargaining with asymmetric information: An empirical study of plea negotiations. *Econometrica* 85(2), 419–452.
- Tuttle, C. (2022). Online appendix for ‘racial disparities in federal sentencing: Evidence from drug mandatory minimums’. [https://codytuttle.github.io/tuttle\\_mandatory\\_minimums\\_online\\_appendix\\_2022.pdf](https://codytuttle.github.io/tuttle_mandatory_minimums_online_appendix_2022.pdf).
- Tuttle, C. (2023). Racial disparities in federal sentencing: Evidence from drug mandatory minimums. Working paper, University of Texas.
- United States Department of Justice (2018). Justice manual. Technical report, Washington, DC.
- USSC (2019). First step act. [https://www.ussc.gov/sites/default/files/pdf/training/newsletters/2019-special\\_FIRST-STEP-Act.pdf](https://www.ussc.gov/sites/default/files/pdf/training/newsletters/2019-special_FIRST-STEP-Act.pdf).

## A Appendix

**Table A.1:** Drug Data - Summary Statistics

	(1) All Cases	(2) Black	(3) Hispanic	(4) White
<i>Drug Type:</i>				
cocaine	0.264	0.253	0.341	0.113
crack	0.181	0.461	0.0395	0.0453
heroin	0.142	0.144	0.164	0.0879
marijuana	0.172	0.0459	0.283	0.134
meth	0.241	0.0952	0.172	0.620
<i>Defendant Characteristics:</i>				
criminal history points	4.112	5.698	2.394	5.330
female	0.124	0.0706	0.117	0.221
high school	0.569	0.661	0.423	0.739
age	35.43	35.91	34.04	37.68
<i>Outcomes:</i>				
gun involved	0.101	0.144	0.0595	0.122
trial	0.0295	0.0452	0.0233	0.0183
# of drug types charged	1.306	1.508	1.178	1.262
percent weight	0.736	0.669	0.753	0.805
conspiracy	0.479	0.501	0.466	0.474
MM imposed	0.219	0.271	0.200	0.177
safety valve	0.314	0.150	0.465	0.245
sentence length	72.59	93.85	58.84	68.82
Observations	42437	14122	19373	8942

*Notes:* Figures represent means. Sample includes all cases with weights between 20% and 180% of the threshold weight. All variables are binary except for criminal history points, age, percent weight, and sentence length. Sentence length is measured in months.

**Table A.2:** US Attorney Data - Summary Statistics

	mean	sd
number of US attorneys per district	3.168	(0.905)
months served	48.28	(26.40)
number of cases	134.2	(150.9)
bunching propensity measure	9.750	(12.84)
max bunching propensity within district	16.41	(15.05)
min bunching propensity within district	2.752	(2.822)
# of Attorneys		282

*Notes:* This table considers characteristics of the US Attorneys data set. Note that number of cases only includes drug trafficking cases with weights between 50% and 150% of the threshold weight; many more cases are prosecuted during an Attorney's tenure. The propensity measures are residualized on district and year, and can be thought of as the proportion of cases during an Attorney's tenure that are charged at the threshold weight.

**Table A.3:** Bunching Propensity Randomization Check

	(1)	(2)	(3)	(4)
	bunch	bunch_score	bunch_score	bunch_score
F-Value:	7.01	1.30	1.10	0.75
F-Test:	0.000	0.237	0.373	0.687
mean:	9.582	9.482	9.507	8.708
Attorney # of Cases	All	All	$\geq 25$	$\geq 50$
R Squared	0.184	0.871	0.905	0.935
N	12216	12216	10927	8553

*Notes:* Here I regress the residualized bunching propensity measure on defendant characteristics omitting fixed effects. Covariates included in the regression are drug type, sex, criminal history points, age and age squared, a binary measure for college, a binary measure for illegal alien, and the proportion of cases with a White defendant for each US Attorney.



**Table A.4:** MM Effects for all Cases Together*Panel A: Varying Controls*

	(1) sentence length	(2) resid. sent. 1	(3) resid. sent. 2	(4) resid. sent. 3
legal effect	10.63	10.78	10.30	9.645
left side 95% CI	[70.36, 75.91]	[-0.510, 4.131]	[0.075, 4.326]	[1.624, 5.668]
right side 95% CI	[81.61, 86.23]	[10.76, 14.82]	[10.87, 14.56]	[11.77, 15.24]
manip. effect	17.01	13.58	7.219	7.054
fit value	70%	70%	70%	70%
baseline controls	no	yes	yes	yes
additional controls	no	no	yes	yes
fixed effects	no	no	no	yes
N of obs fit on	33691	33691	33691	33691

*Panel B: Varying Fit Values*

	(1) sentence length	(2) sentence length	(3) sentence length	(4) sentence length
legal effect	9.502	10.98	9.832	6.676
left side 95% CI	[71.36, 77.15]	[70.63, 77.22]	[69.34, 76.24]	[74.68, 79.39]
right side 95% CI	[81.61, 86.23]	[81.61, 86.23]	[81.61, 86.23]	[81.61, 86.23]
manip. effect	17.01	17.01	17.01	17.01
fit value	60%-80%	64%	62%	79%
N of obs fit on	-	32177	31706	35067

*Notes:* This table gives the regression discontinuity results for the full sample of cases. Panel A gives the preferred specification in column 1, which is the simple linear case fitted at 70% with no controls. Columns 2, 3, and 4 then add various controls (through residualization), still fitting at 70%. Baseline controls include criminal history points, drug type, and racial group. Additional controls include defendant age, defendant education, whether a gun was used, whether the defendant is female, and whether the case went to trial. Fixed effects are at the district and year level. Panel B repeats the analysis in column 1, but now using various fitting schemes. Column 1 averages all fits between 60% and 80%, while columns 2-4 give the median, upper bound, and lower bound fits, respectively. Confidence intervals are determined using the standard error of the expected prediction.

**Table A.5:** Prosecutor analysis - racial disparity in sentence length

	(1)	(2)	(3)	(4)	(5)	(6)
Black	14.35*** (2.698)	13.73*** (2.954)	14.35*** (3.684)	15.18*** (2.451)	14.37*** (2.520)	12.84*** (3.026)
Hispanic	4.454*** (1.577)	4.919*** (1.758)	6.063*** (2.065)	6.080*** (1.610)	6.433*** (1.799)	6.416*** (2.094)
bunching propensity	-0.0807 (1.033)	0.548 (1.244)	1.179 (1.245)	0.226 (1.111)	0.847 (1.361)	1.232 (1.393)
Black*bunching propensity	0.424** (0.163)	0.411** (0.184)	0.400** (0.187)	0.0829 (0.162)	0.0965 (0.160)	0.455** (0.182)
Hispanic*bunching propensity	0.321*** (0.0817)	0.283*** (0.0935)	0.385*** (0.140)	-0.0424 (0.113)	-0.0498 (0.126)	0.200 (0.176)
Includes Bunching Point	yes	yes	yes	no	no	no
Attorney # of Cases	All	$\geq 25$	$\geq 50$	All	$\geq 25$	$\geq 50$
R Squared	0.309	0.314	0.316	0.347	0.353	0.355
N	12210	10927	8553	11040	9889	7825

*Notes:* The dependant variable is sentence length at disposition. All specifications include the full set of controls; defendant characteristics, the percent of cases with a White defendant per US Attorney, and time and district fixed effects. The first three specifications include the bunching point while columns 4-6 omit all cases at the bunching point. This tests whether disparities in sentence length between races is driven by drug weight manipulation. Standard errors are clustered at the district level.

Significance levels: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

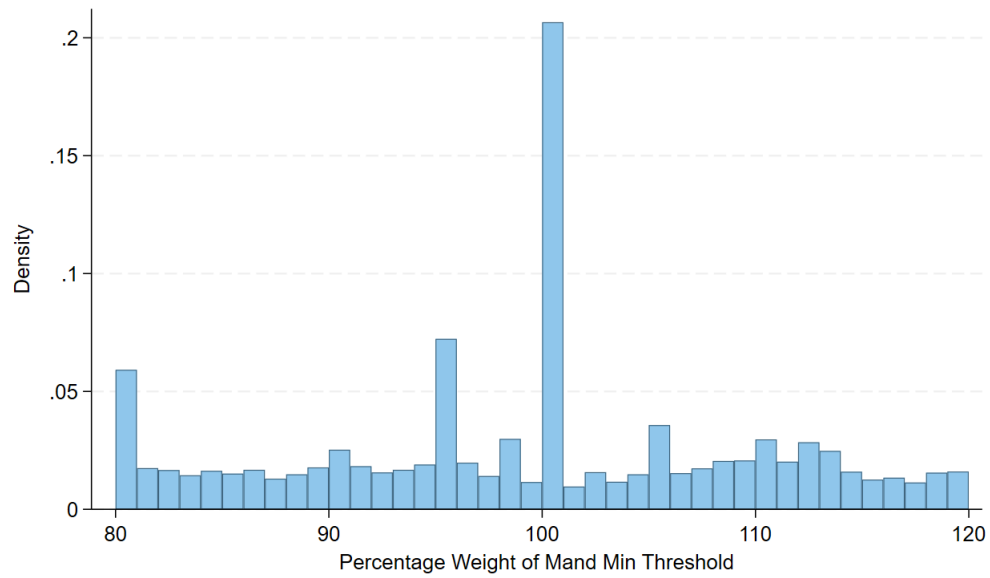
**Table A.6:** Prosecutor Analysis - Omitting High and Low Case-per-month Attorneys

	(1) bunch	(2) sent length	(3) bunch	(4) sent length
Black	-0.0244* (0.0123)	14.82*** (2.857)	-0.0194 (0.0124)	14.48*** (2.569)
Hispanic	-0.0194** (0.00961)	4.270** (1.900)	-0.0171* (0.00932)	4.525** (1.895)
bunch_score	-0.00119 (0.00146)	-0.292* (0.148)	0.000389 (0.00147)	-0.172 (0.115)
Black*bunch_score	0.00369*** (0.00116)	0.303** (0.138)	0.00296** (0.00123)	0.276** (0.120)
Hispanic*bunch_score	0.00156** (0.000769)	0.294*** (0.0942)	0.00115 (0.000767)	0.213** (0.0857)
Attorney Cases per Month	$\geq 0.5$	$\geq 0.5$	$\leq 9$	$\leq 9$
R Squared	0.191	0.316	0.192	0.311
N	11190	11190	10701	10701

*Notes:* All specifications include the full set of controls; defendant characteristics and time and district fixed effects. The first two specifications omit US Attorney spell cases with the bottom 10 percent of cases per month. Columns 3 and 4 omit the top 10 percent. Note that this is different than the sample restrictions in the main specifications - there I omit attorneys who see very few cases total which removes several attorneys who only served for a short amount of time. In this analysis I omit attorneys on a cases-per-month basis which may keep short-term attorneys if they fall within the assigned ranges.

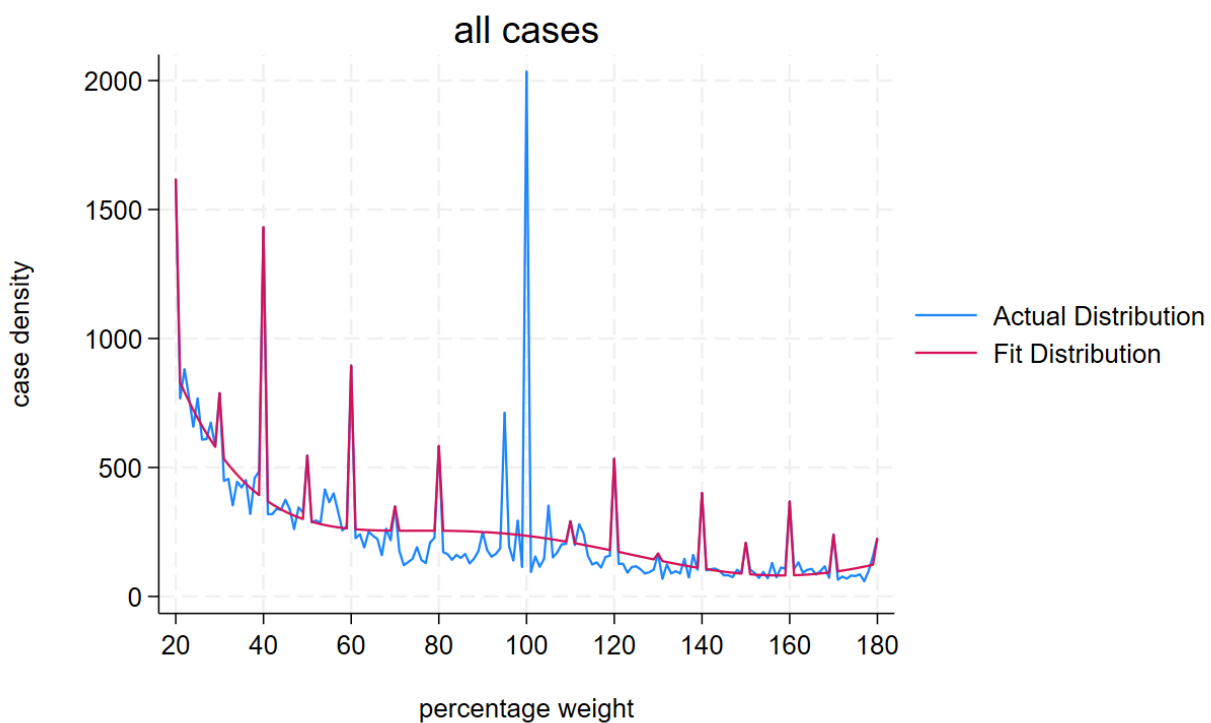
Significance levels: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

**Figure A.1:** Distribution of all cases near the threshold



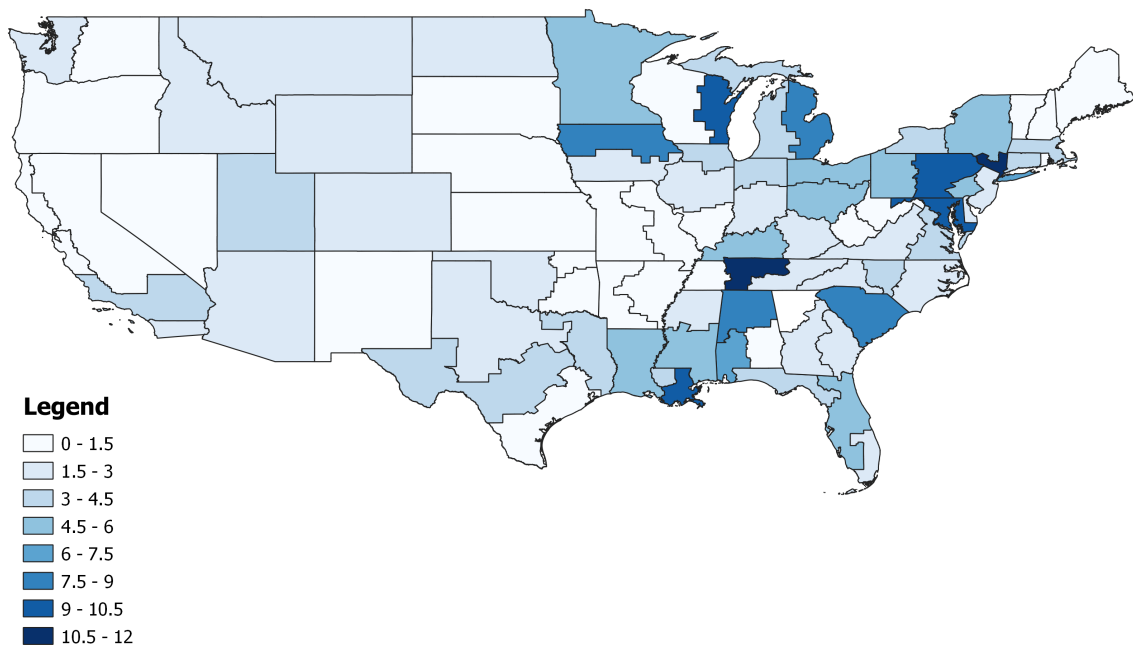
*Notes:* This histogram shows bunching of cases at the 10-year mandatory minimum weight threshold, indicating the presence of case manipulation.

**Figure A.2:** Distribution of cases with fit



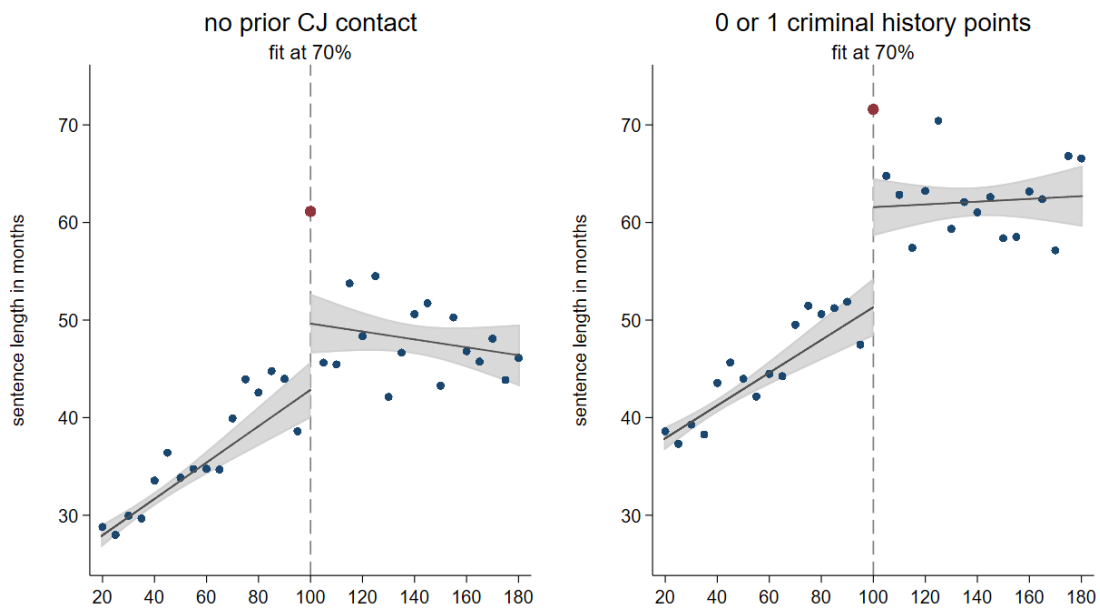
*Notes:* This figure considers which cases are being manipulated to the bunching point. I plot the distribution of cases and then fit a 5<sup>th</sup> order polynomial controlling for internal bunching that occurs at round points (I control for every 10% value). Missing mass is identified as areas where cases are below the fit polynomial - this appears to be primarily from weights at 70% to 95%.

**Figure A.3:** Percent of Cases Bunched by District



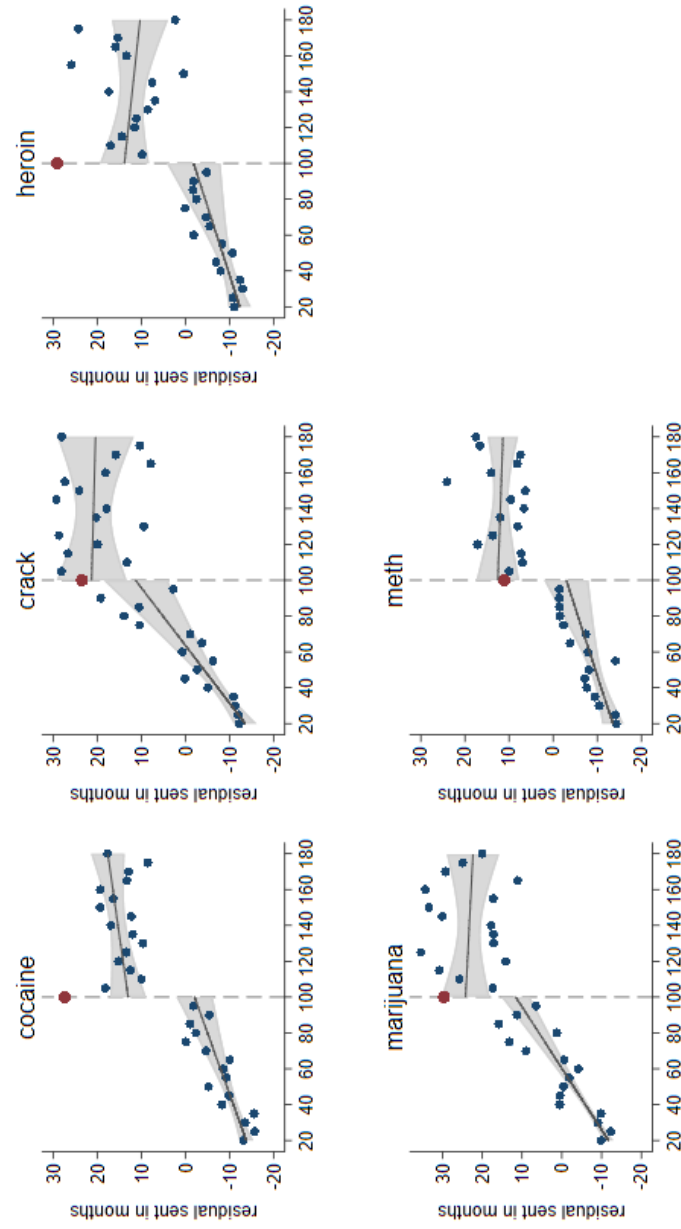
*Notes:* This figure shows variation in bunching propensity across district. It presents the percent of cases bunched for all cases before December of 2018 across each district.

**Figure A.4:** Main results by criminal history group



*Notes:* These are the main regression discontinuity results for cases in the first two criminal history categories, generally considered low-history defendants. Group 1 includes only defendants who have had no previous encounters with the criminal justice system, including events that would lead to zero criminal history points, such as arrest. Group 2 includes individuals who have no points but have had some encounters with the justice system, and individuals with one point. Both discontinuities are fit using the 70% cutoff for extrapolation and use linear fits.

**Figure A.5:** Main results by drug type



*Notes:* This figure illustrates the regression discontinuity design controlling for race and criminal history points, fit at 70% for each drug type. Data at the bunching point is larger and with a different color simply to emphasize differences in manipulation effects between drug types.