Online Appendix for "How Much do Mandatory Minimums Matter?"

Spencer Cooper - 2025

Contents

A Mechanism and Treatment Intensity - US Attorney Analysis

The following is a supplemental analysis that considers heterogeneity in manipulation across US Attorney spells. The basic idea is that US Attorney administration turnover may create variation in the propensity of bunching within a district. While secondary to the main paper, the analysis provides further validity of the main results and additional context surrounding MM sentencing and drug weight manipulation. Specifically, this analysis serves two main purposes: (1) it gives additional evidence that prosecutors drive bunching and that bunching leads to higher sentences at the threshold, (2) it considers whether higher bunching administrations drive race disparities.

Data

This analysis combines the primary drug dataset with a unique, hand collected data set of US Attorneys from years 2013 to 2020 to consider heterogeneity by treatment intensity. 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.[1] I match this data with the USSC drug data to analyze selection patterns for cases between June 2013 and September 2020. This gives a sample of 22,606 cases.

Table A.9 gives a few key statistics about US Attorney administrations. For the study period of 2013 to 2020, each district had an average of 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.

US Attorney manipulation measure

I utilize the variation in bunching propensity by US Attorney administration as a pseudo-random measure of the probability of exposure to drug weight manipulation. The idea is that non-bunching and low-bunching administrations serve as reliable counterfactual distributions compared against high-bunching administrations. This is a similar strategy as described in Frandsen (2017) and practiced in Goncalves and Mello (2021). However, the Frandsen (2017) method relies on comparing observation counts for values neighboring and at the bunching point to identify bunching and non-bunching agents. Because of round number bias, this method does not work well in my setting - an attorney who practices little or no manipulation may still exhibit higher observation counts at the bunching point compared to neighboring values.

[1]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

Instead, I construct a continuous, residualized bunching propensity measure. 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. I then simply divide attorney spells into halves based on this bunching score and compare outcomes across these two groups. The bunching propensity score is calculated as the proportion of cases at the bunching point within an attorney spell, residualized on district fixed effects. Residualizing on district fixed effects accounts for geographic differences in case volumes and drug weight distributions and allows the bunching measure to capture differences in manipulation relative to other spells within a district. Figure A.2 highlights the variation in bunching by district.

One might worry that differences between the high-bunching and low-bunching Attorney groups are not measuring differences in manipulation decisions, but are the result of changes in sample composition or criminal behavior. That is, I need that 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 activity or law enforcement, I can broadly test for selection on observables by regressing bunching propensity on defendant characteristics. Table A.10 reports F-Statistics and tests for regressions run on observables. The joint F-test for the simple bunching binary measure returns an F-value of 7.11 and a p-value of 0.000, indicating non-random selection of who has their case manipulated. However, when using the residualized attorney bunching propensity score, the resulting F-statistic decreases significantly to a value of 1.26, indicating conditional random assignment to high or low bunching US Attorney spells. Limiting the sample to attorneys who see more than 25 or 50 drug cases within the weight range yields similar results. 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.

If US Attorney spells really create variation in manipulation, bunching levels should change with US Attorney turnover. To investigate this, I create an event study for when a US Attorney changes and the new Attorney has a higher bunching propensity than the previous one within their district. I consider 12-months before and after the US Attorney change, meaning if the change resulted in a US Attorney that served for fewer than 12-months, they are not counted as treated here. Figure A.8 displays the event study graph. Note that following a change to an attorney that bunches more, the proportion of cases at the threshold weight nearly doubles.

Heterogeneity of main results

I now check how both the causal effects and selection patterns differ across the high and low bunching Attorney groups. I first consider the event study format again but now see whether selection patterns change with Attorney turnover. The expected result here is not completely obvious; US Attorney spells with low bunching may still exhibit sentence lengths above the counterfactual if round number bias is low. However, if high bunching is indicative of case manipulation, the expected impact of US Attorney turnover is a stronger selection impacts. Likewise, the main results suggest these larger selection patterns should be driven by minority cases.

Figure A.9 displays the result. Note that selection impacts are only measured for cases at the threshold weight. Thus, by reducing the sample to cases at the threshold within a year of a qualifying US Attorney change, the number of observations become thin, resulting in noisy results. Even still, the graph shows a clear change in trend for the non-White groups with increases in selection impacts following the US Attorney change and no real change for White defendants. This provides evidence that bunching and observed selection impacts are at least in part driven by prosecutor decisions.

I then run the main discontinuity analysis again split by US Attorney spell groups. Specifically, I compare effects across the split of the top half of bunching attorney spells versus the bottom half. Note that this analysis still has smaller sample sizes than the main analysis due to data constraints from the US Attorney data, which only includes years 2013-2020. In concordance with the event study, I find that the top half of spells exhibit much larger selection impacts than the bottom half, with the high-bunch group averaging gaps between 11.24 and 22.08 months and the low-bunch group averaging gaps between 4.55 and 9.94 months. However, the causal effects are on average slightly larger for low-bunching spells. The disparity in discontinuities is considerably smaller than in selection, but is consistent across specifications. These results are displayed in Table A.11.

B Additional tables

Table A.1: Drug data - summary statistics

	(1) All Cases	(2) Black	(3) Hispanic	(4) White
Drug type:				
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
Defendant characteristics:				
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
Outcomes:				
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: MM effects for all cases together

	(1) sent length	(2) resid sent 1	(3) resid sent 2	(4) resid sent 3
discontinuity	10.63	11.08	10.10	9.65
left side 95% CI	[70.36, 75.91]	[-0.61, 4.00]	[1.17, 5.58]	[1.62, 5.67]
right side 95% CI	[81.61, 86.23]	[10.96, 14.98]	[11.78, 15.61]	[11.77, 15.24]
selection impact	17.01	12.40	11.24	7.05
fit value	70%	70%	70%	70%
baseline controls	no	yes	yes	yes
fixed effects	no	no	yes	yes
pros. decision controls	no	no	no	yes
N of obs fit on	33691	33691	33691	33691

Notes: This table gives the regression discontinuity results for the full sample of cases. Column 1 presents 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, defendant age, defendant sex, and defendant education. Fixed effects are at the district and year level. Prosecutor decision controls include whether a gun was used in the offense, the number of drug types charged in the case, and whether the case went to trial.

Table A.3: MM effects with varying fits

	(1) sent length	(2) sent length	(3) sent length	(4) sent length
discontinuity	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]
selection gap	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 repeats the simple linear discontinuity displayed in Table 2, but now uses 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.4: Main effects by drug type comparison

	(1) cocaine	(2) crack	(3) heroin	(4) marijuana	(5) meth
discontinuity	15.00	9.873	15.65	12.51	15.51
left side 95% CI	[-6.26, 1.90]	[3.86, 18.28]	[-8.02, 4.08]	[7.87, 14.99]	[-8.05, 1.89]
right side 95% CI	[8.98, 17.02]	[13.42, 29.02]	[8.26, 19.30]	[18.19, 30.16]	[7.74, 17.33]
selection gap	14.43	2.181	15.19	5.430	-1.474
fitted mean at 99%	-0.157	12.56	-1.276	14.57	-1.716
fit value	70%	70%	70%	70%	70%
N of obs fit on	5865	5233	3035	5256	4272

Notes: The dependent variable is sentence length residualized against criminal history points and race. 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 causal effect is the regression discontinuity between the two extrapolated fits at the 10-year MM cutoff. Evidence of selection is seen in 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.

Table A.5: Main effects by criminal history and safety valve

	(1) no prior history	(2) low history	(3) no safety valve	(4) yes safety valve
discontinuity	6.708	10.46	16.11	5.397
left side 95% CI	[1.41, 7.12]	[-2.16, 3.53]	[-3.79, 2.88]	[0.84, 3.43]
right side 95% CI	[8.11, 14.16]	[8.38, 14.25]	[13.01, 18.69]	[6.30, 9.04]
bunch effect	9.830	11.21	11.93	-5.984
fitted mean at 99%	4.26	0.68	2.14	2.61
fit value	70%	70%	70%	70%
N of obs fit on	3395	6828	16314	7197

Notes: The dependent variable is sentence length residualized against drug type and race. 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 causal effect is the regression discontinuity between the two extrapolated fits at the 10-year MM cutoff. Evidence of manipulation is seen in 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.

Table A.6: High vs low bunching districts

	(1) sent length	(2) resid sent 1	(3) resid sent 2	(4) resid sent 3
Panel A: ≤ 3% bunching				
discontinuity	9.674	9.387	9.426	9.465
N of obs fit on	25759	25759	25759	4812
Panel B: ≤ 4% bunching				
discontinuity	11.06	11.31	10.82	10.92
N of obs fit on	33390	33390	33390	33390
Panel C: ≤ 5% bunching				
discontinuity	10.98	11.34	10.53	10.63
N of obs fit on	37140	37140	37140	4812
Panel D: ≥ 5% bunching				
discontinuity	9.467	7.862	6.505	5.734
N of obs fit on	12784	12784	12784	12784
fit value	70%	70%	70%	70%
baseline controls	no	yes	yes	yes
fixed effects	no	no	yes	yes
pros. decision controls	no	no	no	yes

Notes: This table presents the main discontinuity effects for sub-samples of districts based on the proportion of cases at the bunching point. Panel A, B, and C give districts with low amounts of bunching while Panel D gives districts where more than 5% of all cases were at the bunching point. CI's are omitted for space, but are similar to mainline results. All extrapolation is fitting between 20% and 70% for the left side of the threshold and 105% and 180% on the right side.

Table A.7: Discontinuity analysis with multiple drug weight controls

	(1) All cases	(2) Black	(3) Hispanic	(4) White
discontinuity	9.520	11.86	10.18	11.02
left side 95% CI	[1.63, 5.64]	[2.60, 11.07]	[0.54, 4.97]	[-6.01, 2.01]
right side 95% CI	[11.65, 15.11]	[14.72, 23.23]	[10.89, 15.16]	[6.09, 12.42]
selection gap	5.573	3.665	3.856	-2.249
fit value	70%	70%	70%	70%
N of obs on fit	33691	11205	15296	7190

Notes: The dependant variable is residual sentence length. All specifications include controls for defendant characteristics, time and district fixed effects, and the new other drug type weight controls. These are weight as a percent of the 10-year mandatory minimum threshold for up to 4 other drug types other than the primary type. Each specification is fit at 70%. Selection gap gives the difference between the right fitted regression and the actual sentence length at the threshold.

Table A.8: Discontinuity analysis with weight range observations included

	(1) sentence length	(2) resid sent 1	(3) resid sent 2	(4) resid sent 3
Panel A: Minimum of weight range				
discontinuity	9.467	13.00	9.192	11.02
left side 95% CI	[71.70, 76.97]	[-4.43, 0.11]	[-5.05, -0.84]	[-1.84, 2.19]
right side 95% CI	[81.55, 86.20]	[8.86, 12.96]	[4.37, 8.09]	[9.52, 13.03]
	{33911}	{33911}	{33911}	{33911}
selection gap	15.20	7.156	12.38	5.319
Panel B: Median of weight range				
discontinuity	13.69	11.68	10.91	11.09
left side 95% CI	[68.19, 73.37]	[-1.40, 3.077]	[-5.90, -1.76]	[-0.67, 3.26]
right side 95% CI	[82.19, 86.81]	[10.56, 14.64]	[5.15, 8.86]	[10.71, 14.21]
	{33546}	{33546}	{33546}	{33546}
selection gap	16.02	14.79	5.372	8.309
Panel C: Maximum of weight range				
discontinuity	13.22	16.93	9.799	12.06
left side 95% CI	[67.44, 72.39]	[-6.69, -2.42]	[-6.32, -2.40]	[-1.78, 1.94]
right side 95% CI	[80.86, 85.51]	[10.37, 14.47]	[3.52, 7.25]	[10.44, 13.97]
	{33546}	{33546}	{33546}	{33546}
selection gap	11.37	6.788	6.706	3.471
fit value	70%	70%	70%	70%
baseline controls	no	yes	yes	yes
additional controls	no	no	yes	yes
fixed effects	no	no	no	yes

Notes: This table gives the regression discontinuity results for the full sample of cases including cases with imprecise weight measures. For each measure, I use the precise weight for cases where it is available and then vary how the range measure is considered. In Panel A, I use the minimum value of the range as the drug weight measure. In Panel B, I use the median of the drug weight range. And in Panel C, I use the maximum weight in the drug range. The four specifications use the same control schemes as used in the main analysis. All regressions are fit at 70% of the threshold weight and use a linear fit. Sample size is given in curly braces.

Table A.9: US attorney data - summary statistics

	mean	sd
number of US attorneys per district	3.125	(0.894)
months served	48.62	(25.94)
number of cases	343.2	(390.4)
percent of cases bunched	3.130	(3.594)
max bunch pct within district	5.681	(7.559)
min bunch pct within district	1.299	(2.001)
# 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 20% and 180% of the threshold weight; many more cases are prosecuted during an Attorney's tenure.

Table A.10: Bunching propensity randomization check

	(1) bunch	(2) bunch_score	(3) bunch_score	(4) bunch_score
F-Value:	7.11	1.26	0.90	0.94
F-Test:	0.000	0.263	0.544	0.507
Attorney # of Cases	All	All	≥ 25	≥ 50
N	22606	22606	20803	16986

Notes: Here I regress the residualized bunching propensity measure on defendant characteristics omitting district 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.11: High vs low bunching US Attorney spell

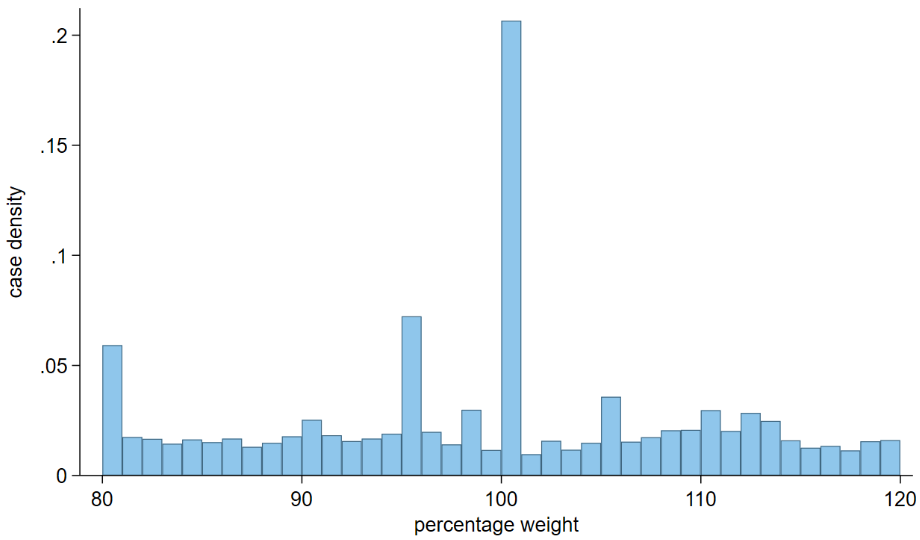
	(1) sent length	(2) resid sent 1	(3) resid sent 2	(4) resid sent 3
Panel A: Low-bunch				
discontinuity	15.71	15.11	13.49	12.08
left side 95% CI	[62.57, 74.11]	[-8.50, 2.09]	[-7.01, 3.24]	[-4.48, 4.91]
right side 95% CI	[79.70, 88.62]	[7.85, 16.21]	[7.80, 15.70]	[8.84, 16.06]
selection gap	7.490	8.267	9.938	4.550
N of obs fit on	4812	4812	4812	4812
Panel B: High-bunch				
discontinuity	10.23	9.875	9.088	8.918
left side 95% CI	[63.78, 72.41]	[-4.65, 3.28]	[-3.37, 4.36]	[-2.47, 4.65]
right side 95% CI	[74.72, 82.28]	[5.88, 12.92]	[6.39, 13.20]	[7.11, 13.34]
selection gap	22.08	17.68	16.16	11.24
N of obs fit on	7393	7393	7393	7393
fit value	70%	70%	70%	70%
baseline controls	no	yes	yes	yes
fixed effects	no	no	yes	yes
pros. decision controls	no	no	no	yes

Notes: This table presents the main discontinuity effects again, now split by US Attorney bunching groups. Panel A gives the discontinuity estimates for the low-bunching Attorney spells while Panel B reports effects for high-bunching spells. All extrapolation is fitting between 20% and 70% for the left side of the threshold and 105% and 180% on the right side.

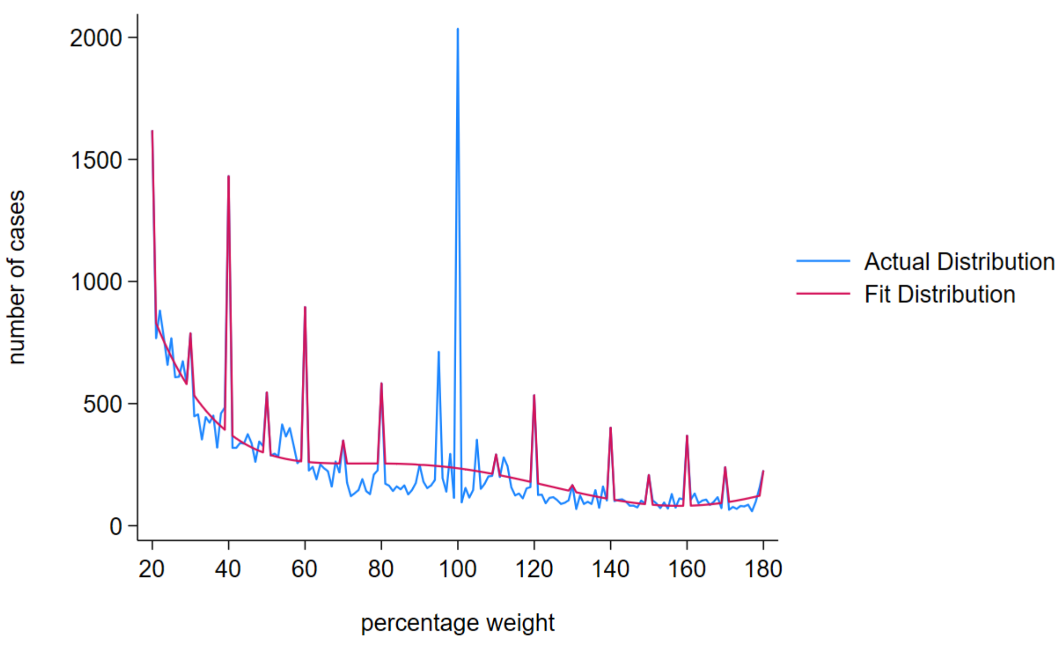
C Additional figures

Figure A.1: Case distribution by sentencing weight

(a) Distribution of all cases near the threshold

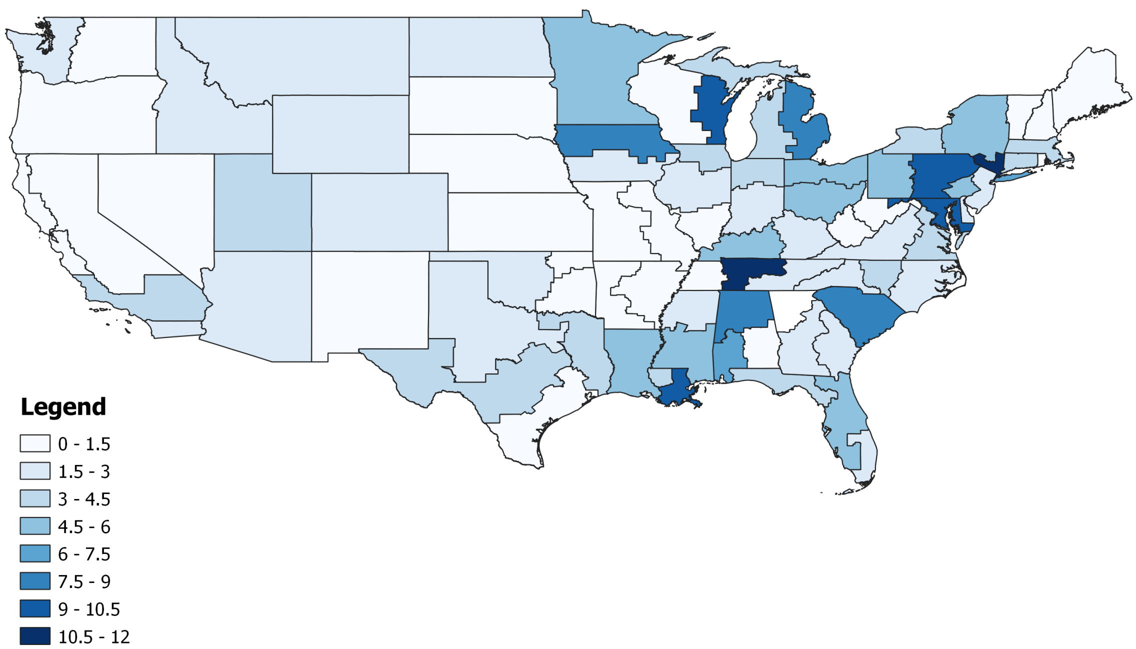


(b) Distribution of cases with fit



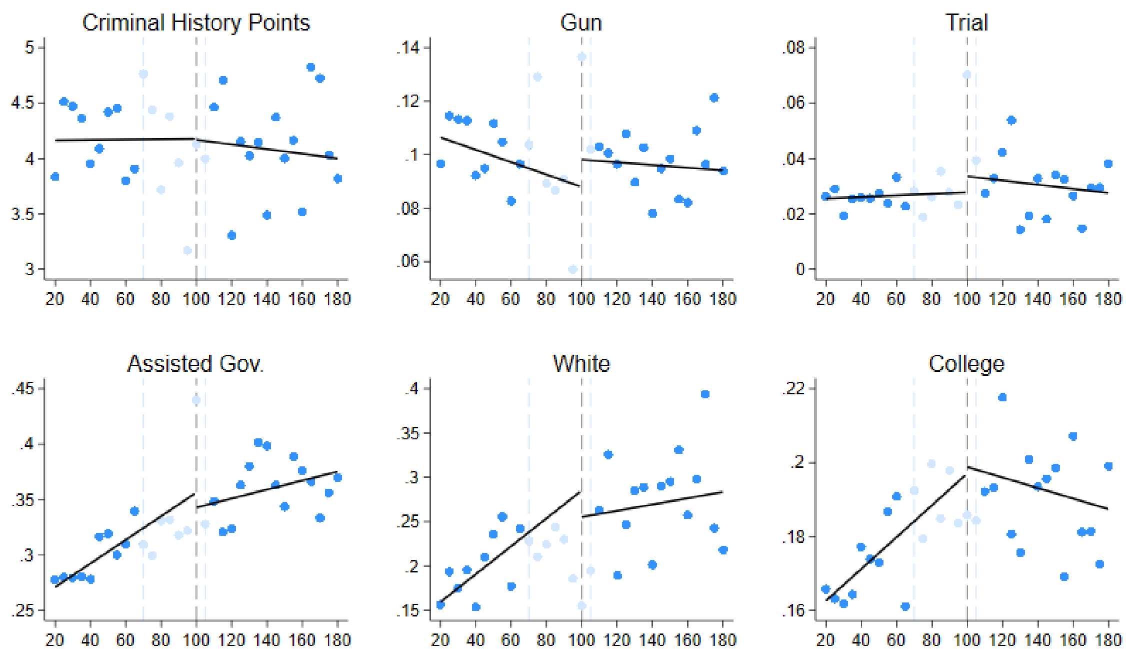
Notes: This figure considers which cases are being manipulated to the bunching point. I plot the distribution of cases and then fit a 5^{th} order polynomial controlling for internal bunching that occurs at round points (I control for every 10% value). Missing mass is identified as areas where cases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

Figure A.2: Percent of cases bunched by district



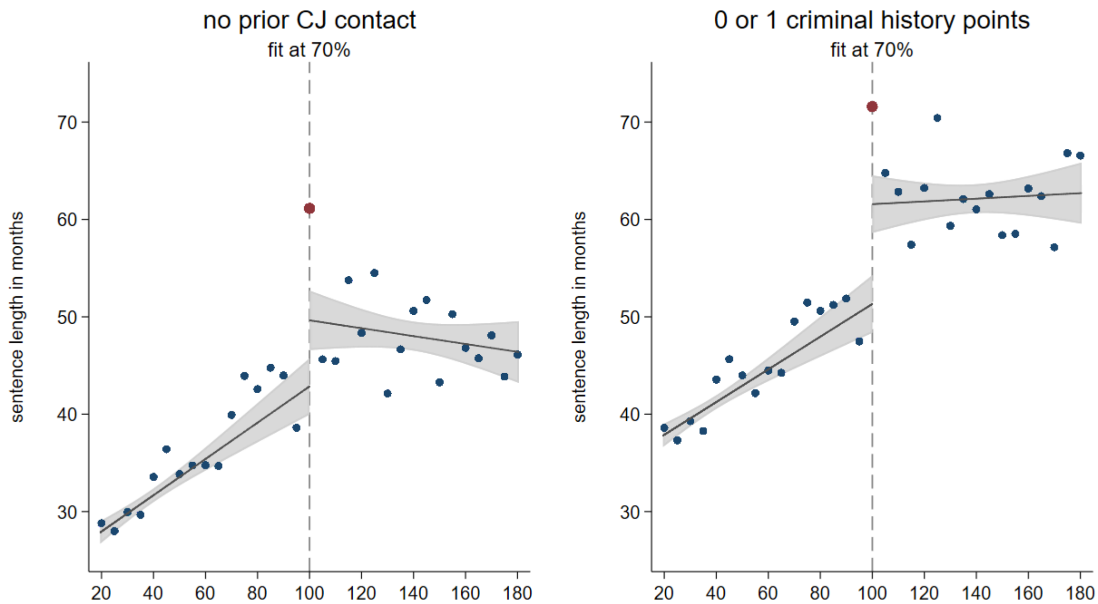
Notes: This figure shows variation in bunching propensity across district. It presents the percent of cases bunched for all cases before December of 2018 across each district.

Figure A.3: Smoothness Test



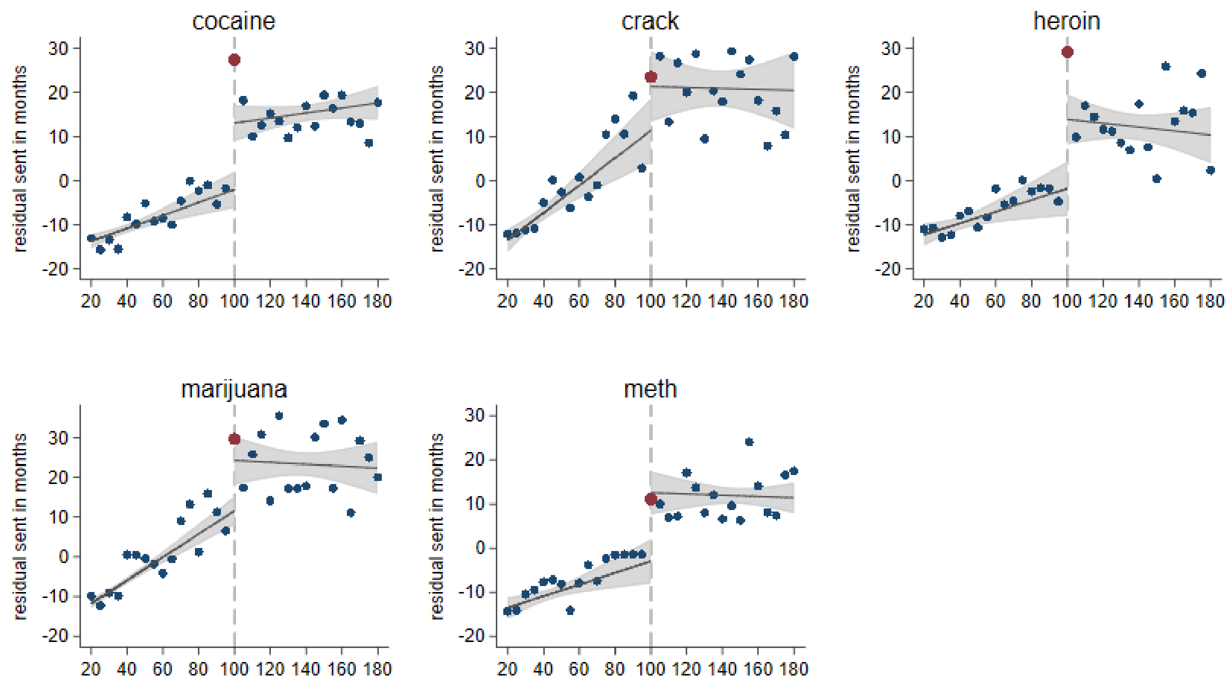
Notes: This figure presents regression discontinuity analyses for a range of factors, determinants, and outcomes related to sentence length. The graphs depict raw discontinuities using a donut design, with cutoffs at 70% and 105% of the threshold weight. Data points outside the fitting bandwidth are displayed with reduced opacity to highlight the observations used for local polynomial estimation.

Figure A.4: Main results by criminal history group



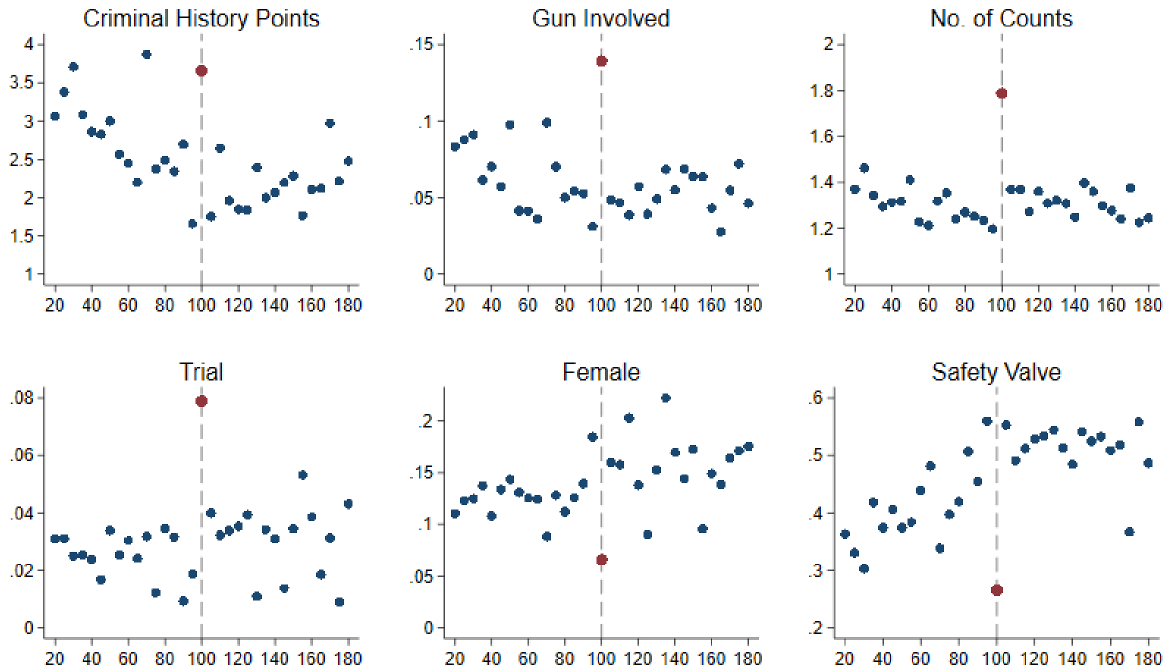
Notes: These are the main regression discontinuity results for cases in the first two criminal history categories, generally considered low-history defendants. Group 1 includes only defendants who have had no previous encounters with the criminal justice system, including events that would lead to zero 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



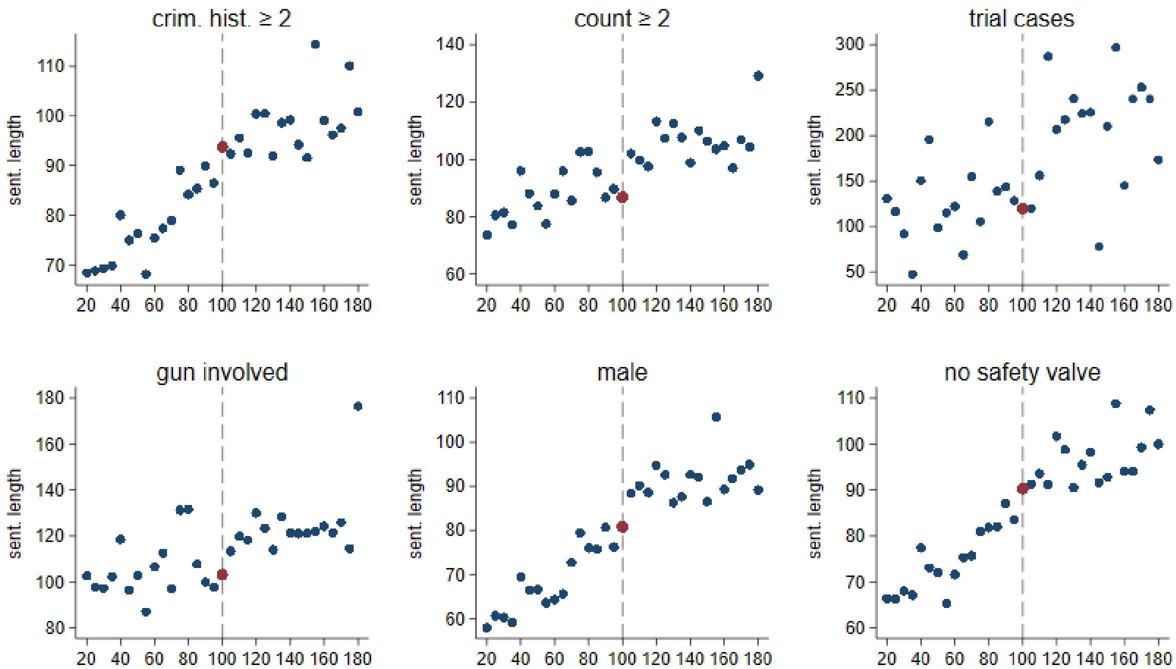
Notes: This figure illustrates the regression discontinuity design controlling for race and criminal history points, fit at 70% for each drug type. Data at the bunching point is larger and with a different color simply to emphasize differences in selection patterns between drug types.

Figure A.6: Factors being selected on



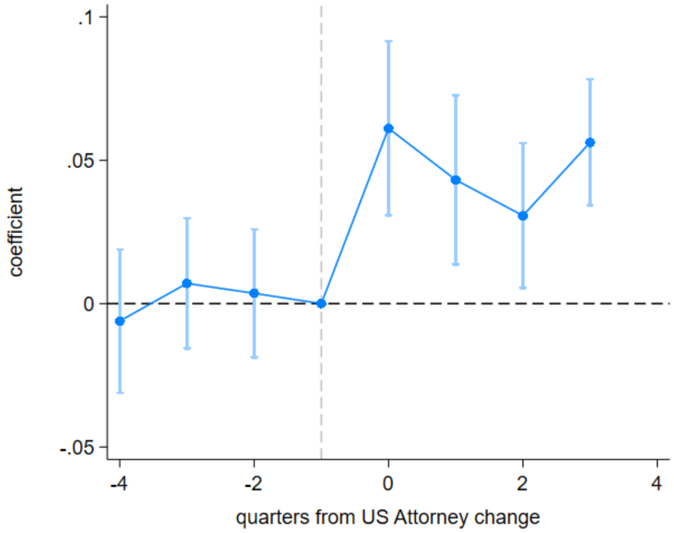
Notes: The following shows selection patterns for minority defendants in cocaine or heroin cases. Each graph displays a different key variable measured across drug weight as a percentage of the threshold weight. These are shown in 5% bins, except at the threshold weight which only contains cases charged 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 red simply for emphasis.

Figure A.7: Factors being selected on



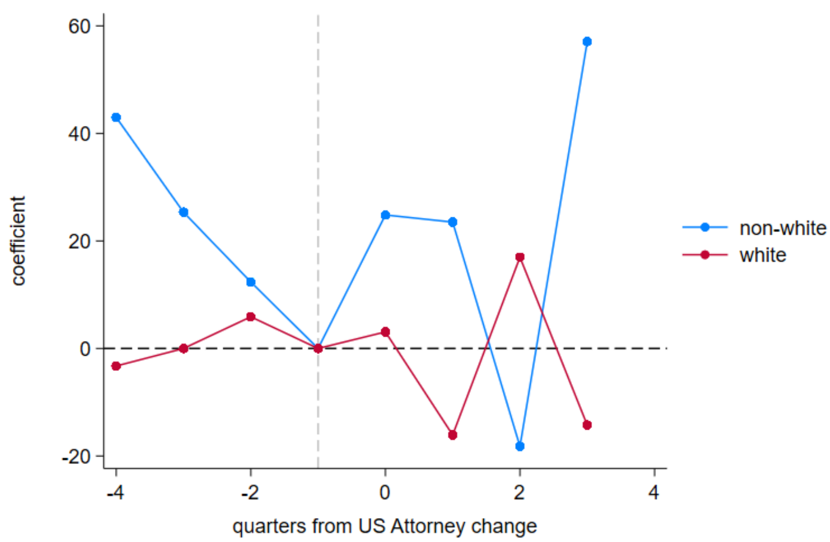
Notes: The following gives sentence length over drug weight for specific subsets of White defendants. Each graph represents a sample restricted to cases more likely to experience manipulation across some correlate. Note that graph titles reflect how sample is restricted rather than a y-variable. In all cases, the dependent variable is sentence length in months.

Figure A.8: Bunching by US Attorney turnover



Notes: Each point gives the proportion of cases bunched within a district in quarterly time relative to the US Attorney turnover. This is specific to cases where a district changes from a US Attorney that bunches more than the previous Attorney(s).

Figure A.9: Change in Selection Gap at US Attorney turnover



Notes: This figure follows the same design as the event study in Figure A.8, now measuring gaps between observed and predicted sentence lengths. No controls or fixed effects are included due to thin data. Each point gives the average difference between sentence length and counterfactual sentence length using the fitted values of the right-hand side extrapolation.

Online Appendix 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.

Bhuller, M., G. B. Dahl, K. V. Loken, and M. Magne (2020). Incarceration, recidivism and employment. *Journal of Political Economy 128*(4).

Di Tella, R. and E. Schargrodsky (2013). Criminal recidivism after prison and electronic monitoring. *Journal of Political Economy 121*(1), 28–73.

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.

Frandsen, B. R. (2017). *Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design when the Running Variable is Discrete.* Bingly, UK: Emerald Publishing Limited.

Goncalves, F. and S. Mello (2021). A few bad apples? racial bias in policing. *American Economic Review 111*(5), 1406–41.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review 96*(3), 863–76.

Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. `https://sites.lsa.umich.edu/mgms/wp-content/uploads/sites/283/2015/09/incar.pdf`.