Bail Reform & Recidivism Series

Technical Supplement: Quasi-Experimental Methods for Estimating the Impact of Bail Reform on Recidivism in New York City

March 2023

René Ropac with Michael Rempel

DATA COLLABORATIVE FOR JUSTICE AT JOHN JAY COLLEGE
STRENGTH IN NUMBERS

JOHN JAY COLLEGE OF CRIMINAL JUSTICE

Table of Contents

Main Research Report Available Here at Report Landing Page.

The Data Collaborative for Justice (DCJ) at John Jay College of Criminal Justice houses a group of research initiatives that raise important questions and share critical research about the criminal justice system and its role in creating safe, just, and equitable communities. DCJ conducts data analysis and research on enforcement in the community, the adjudication of cases in the courts, and the use of confinement in jails and prisons. DCJ's work has informed policy reforms, facilitated partnerships between researchers and government agencies across the country, spurred new scholarly research on lower-level enforcement, and has been cited extensively in the press. For more information about the Data Collaborative for Justice please visit: https://datacollaborativeforjustice.org/

Section 1 - Analytic Strategy for Producing Unbiased Estimates of Bail Reform's Impact on Recidivism

This technical supplement provides a comprehensive description of the **propensity score** and **inverse probability of treatment weighting (IPTW)** methods used in this study to produce recidivism estimates based on bail reform and comparison samples that were statistically similar at baseline.

All methods were applied using data from the New York State Office of Court Administration (OCA).[1] The OCA data includes all cases filed in criminal court, omitting the fraction of cases that the prosecutor declines. Hence this is technically a study of *prosecuted arrests and re-arrests*.

As discussed in the main report, all samples were drawn either from the first half of 2019 (pre-reform) or the first half of 2020 (post-reform). We possessed criminal history data for the two years prior to each sampled individual's instant case; and we tracked recidivism through June 30, 2022, affording from two years to 30 months of recidivism tracking depending on exactly when the instant case was arraigned.

Adjusting for Selection Bias: Propensity Score Modeling and IPTW

Methodological Rationale

Propensity score adjusted models are preferable to traditional regression analysis as they produce estimated treatment effects that are generally closer to the true treatment effects (i.e., the actual recidivism impact of bail reform).[2] First and foremost, propensity score methods help achieve similar distributions of observed baseline characteristics across groups, which reduces the effects of confounding variables and selection bias and yields more valid estimates of the treatment effect on an outcome.[3] The bulk of the work is focused on the development of comparable groups, which mitigates the risk of unintentional bias in model specifications, i.e., the final models only include the main independent variable (i.e., pretrial release status) and the propensity score, which safeguards against p-hacking.[4]

Inverse Probability of Treatment Weighting (IPTW) adds standardized weights to propensity score adjusted models to create a synthetic sample with an increased representation of "rare" cases and a decreased representation of "common" cases in each group.[5] This further facilitates the equal distribution of confounders across groups and reduces the need to trim cases to achieve covariate balance, thereby increasing the external validity of findings (i.e., generalizability to a wide range of cases).

For purposes of the current study, the "treatment" is bail reform, which has the relevant operational effect of leading to the pretrial release of people who would have otherwise faced bail or remand.

Application

We use propensity score adjusted and inverse probability weighted logistic regression models and Cox proportional hazards models to estimate the impact of bail reform on recidivism. Specifically, we compare the likelihood of recidivism between people who had bail set or were remanded to similar people who were released without bail. We also conduct several subgroup analyses to more closely examine the impact of release status on recidivism.[6]

Operational Steps to Adjust for Selection Bias in the Current Study

The creation of the propensity scores and the specification of the inverse probability of treatment weights is an iterative process. At a minimum, the steps include: (1) regressing all plausible predictor variables on outcome variable "treatment" to calculate a propensity score for each observation; (2) creating the inverse probability of treatment weights; and (3) checking whether the two groups are comparable based on all observable characteristics that were used to generate the propensity scores. However, since these steps proved not to be sufficient to create comparable groups in the current study, we also trimmed cases—as an intermediate step between running the propensity score regression model (Step 1) and the creation of the weights (Step 2)—and truncated extreme weights to achieve covariate balance. We recalibrated these procedures several times to create balanced groups while retaining as many observations in the final samples as possible.

The following paragraphs describe in more detail each of **six actual steps** in the current study.

Step 1. Propensity Score Regression Model

To generate a propensity score for each observation, we conducted logistic regression models with the binary outcome variable "treatment" (i.e., arraignment release status; coded 0 = "bail set/remanded" and 1 = "released without bail") and a large number of relevant independent variables, including individuals' demographic factors, current case characteristics, and criminal history variables. (Exhibits T1 through T3 below show the variables included in the propensity score models.) We also included several interaction terms to further reduce covariate imbalance across the groups.[7] Model parsimony and collinearity are of no concern when it comes to the creation of propensity scores.[8]

Step 2. Trimming

We applied asymmetric trimming—that is, we excluded comparison group cases below, and treatment group cases above, a certain propensity score threshold[9]—to two of the three analytic samples (i.e., the mandatory release sample and the bail eligible contemporaneous sample) while no trimming was necessary to achieve covariate balance between the two groups in the bail-eligible pre- vs. post sample. (Please refer to Chapter 2 in the main report for a high level review of the research design and the three analytic samples; and see below for trimming decisions regarding the two analytic samples for which it was necessary.)

To retain as many cases as possible, we started the trimming process conservatively and applied more aggressive trimming as needed. Specifically, we first trimmed comparison group cases that fell below the 5th propensity score percentile in that group and treatment group cases that fell above the 95th propensity score percentile in that group, and checked whether this was sufficient to achieve covariate balance. We found that, for both samples, further trimming was necessary, so we gradually lowered the trimming thresholds several times until the comparison and treatment groups were comparable on observed baseline characteristics.

Trimming for the Mandatory Release Sample

To generate comparable groups in the mandatory release sample, we trimmed comparison group cases below the 26th propensity score percentile and treatment group cases above the 74th propensity score percentile. This resulted in the exclusion of about 25% of observations, reducing the sample size from 21,351 cases to 15,860 cases. Since we excluded a large percentage of the sample, we conducted attrition analysis to better understand the differences between cases that were dropped versus retained.

We found that trimmed observations were disproportionately individuals accused of misdemeanors, individuals with no history of missed court appearances, and individuals with relatively limited or minor criminal records (i.e., individuals without pending cases, no prior arrests, no prior felonies or VFOs, no prior domestic violence cases, and no prior convictions). This pattern is not surprising given the nature of our analysis and New York City judges' release decisions in the year leading up to bail reform. Essentially, the trimming process removed cases for which judges rarely set bail even prior to the implementation of bail reform.

Put in practical terms, release decisions in trimmed cases were generally unaffected by bail reform in the first place—making these cases irrelevant to the fundamental purpose of our study. That is, our analysis estimates the practical impact of bail reform by comparing bail ineligible cases that had bail set or were remanded pre-reform (before they became subject to mandatory release) with similar cases that were released without bail post-reform. However, many cases that became subject to mandatory release post-bail reform were already routinely released pre-reform,[10] which skewed towards people with relatively minor charges and/or no or limited criminal history.[11] Put differently, the original sample included a significant number of bail reform group cases (i.e., cases released post-reform) whose characteristics would have already put them at a high likelihood to be released pre-reform (which is reflected in their high propensity scores), and since the comparison group is only comprised of cases that had bail set or were remanded pre-reform, there was a lack of common support for these treatment group cases.[12]

Trimming for the Bail Eligible Contemporaneous Sample

In the bail-eligible contemporaneous sample, we ultimately trimmed comparison group cases below the 6th propensity score percentile within that group and treatment group cases above the 94th propensity score percentile within that group, which reduced the sample by less than 7%, from 5,985 cases to 5,588 cases. Regarding a practical interpretation of this trimming outcome, it is expected that far less trimming would be necessary among bail-eligible than mandatory release cases, since the former consist largely of violent felonies or other seriously charged cases that, therefore, almost all at times faced bail during the pre-reform period.

Step 3. Application of Weights

We generated *stabilized* inverse probability of treatment weights using the following formulas: for the treatment group, we divided the proportion of individuals in the treatment group by propensity score; for the comparison group, we divided the proportion of individuals in the treatment group by the difference of 1 minus the propensity score. Using stabilized weights rather than non-stabilized weights reduces the variance of the effect estimate.[13] Finally, we standardized the weights by dividing each individual weight by the mean weight (which results in the standardized mean weight equaling 1).[14]

Step 4. Truncation of "Extreme Weights"

Finally, we truncated "extreme" weights (i.e., weights in the bottom and top percentiles of the weight distribution) to further reduce the disproportionate impact of observations with very small or very large weights on the analysis, thereby further decreasing the variance of the effect estimate.[15] However, weight truncation should be done carefully as it also increases the possibility of biased results. Therefore, similar to our trimming approach described above, we initially applied weight truncation conservatively and then progressively increased the percentage of truncated weights to achieve covariate balance (Cole & Hernan, 2008). Specifically, for each sample, we began by truncating weights below the 0.5th percentile and above the 99.5th percentile[16] and incrementally applied more aggressive truncation as needed.

For the mandatory release sample, we truncated weights at the 3rd and 97th percentiles, resulting in a minimum weight of 0.63 and a maximum weight of 1.56.

For the bail-eligible pre vs. post sample, we truncated weights at the 0.5th and 99.5th percentiles, which produced a minimum weight of 0.50 and a maximum weight of 4.72.

Lastly, for the bail-eligible contemporaneous sample, we truncated weights at the 1.5th and 98.5th percentiles, restricting the weights between 0.49 and 3.31.

Step 5. Validation of Covariate Balance

To validate whether we achieved balanced treatment and comparison groups in each sample, we created quintiles based on the propensity scores and ran two-way ANOVAs for each covariate with independent variables *treatment* and interaction term *treatment * quintile*. We considered a covariate to be balanced across groups when neither the effect of treatment nor the effect of the interaction between treatment and quintile was statistically significant.[17] However, because of the relatively large sample sizes in our analyses, we also considered covariates to be balanced when the group differences were consistently below 2 percentage points as well as below a 10% relative difference in each quintile and in the overall sample, even when the ANOVA results were statistically significant.

Propensity score methods allow researchers to mimic randomization on observed covariates. Since proper randomization of individuals to groups can result in statistically significant baseline differences in up to 5% of observed characteristics simply by chance,[18] this principle also applies to the validation of the groups generated by IPTW. Therefore, we considered the samples to be sufficiently comparable when no more than 5% of baseline characteristics were imbalanced based on the criteria outlined above.

Exhibits T1, T2, and T3 show the baseline characteristics of the bail reform group and the comparison group samples before and after IPTW for the mandatory release analysis, the bail-eligible pre vs. post analysis, and the bail-eligible contemporaneous analysis, respectively.

All three original samples were heavily imbalanced on a large number of baseline characteristics, but the final samples are balanced on almost all covariates. In effect, this means that estimates of recidivism impact would have been extremely biased without deploying propensity score and IPTW methods, but that these methods, along with our trimming and truncation strategies, were largely successful in removing such bias. However, as in all quasi-experimental studies, we cannot rule out the existence of bias on *unobservable characteristics* for which we lack data. Chapter 2 of the main report discusses this and other study limitations.

Step 6. Sensitivity Checks

Once we generated balanced samples, we conducted several **sensitivity analyses** to assess the robustness of results. One sensitivity check was to use different propensity score model specifications on the final samples and compare the results. That is, we used a doubly-robust approach by including imbalanced covariates on top of the propensity scores in the regression models to compensate for insufficient covariate balance.[19]

In the bail-eligible pre vs. post sample, we controlled for whether individuals were charged with a felony and whether individuals had any recent domestic violence arrests (these variables are not fully balanced across quintiles in the final sample). This model generated virtually identical results to the primary model that includes propensity score as the only independent variable. Likewise, in the bail-eligible contemporaneous model, we controlled for whether individuals were charged with a felony, whether individuals were charged with a violent felony, and whether individuals had any recent felony convictions, which also yielded highly similar results to the main model. (We did not run doubly-robust models for the mandatory release sample since all covariates are fully balanced.)

Another sensitivity test was to run the propensity score models on several random subsamples. The results based on the subsamples were similar to the results based on the full samples, further confirming the robustness of our findings. We also ran alternative models including propensity scores that were calculated based on specific subsamples. The idea behind this approach is that correctly-specified propensity scores remain valid in subsamples."[20] The results based on the main models (that use the propensity scores calculated based on the full sample) were highly similar to the results based on these alternative models, which further confirms the robustness of our findings.

Lastly, we increased the truncation of the largest weights in the bail-eligible pre vs. post sample, reducing the maximum weight from 4.72 (which is the maximum weight in the main model specification) to 4.00, to ensure that the inclusion of weights above 4.00 does not compromise our results. Reassuringly, the main model and the alternative model with these slightly more aggressively truncated weights created virtually indistinguishable results.

Exhibit T1: Baseline Characteristics of Bail Reform Group and Comparison Group Samples for the Mandatory Release Analysis, Before and After Inverse Probability of Treatment Weighting (IPTW)

	Original Sample		Final Sample	
	Bail Reform Group *(N=16,666)*	**Comparison Group** *(N=4,685)*	**Bail Reform Group** *(N=12,350)*	**Comparison Group** *(N=3,510)*
Demographics				
Age	36.7***	37.6	36.9	36.9
White	11.0%*	12.1%	10.6%	10.6%
Black	48.1%***	53.3%	53.5%	53.5%
Hispanic	34.1%***	31.5%	31.9%	32.1%
Additional Race/Ethnic Groups	6.8%***	3.2%	4.0%	3.9%
Female	19.3%***	8.3%	8.6%	8.9%
Arraignment Borough				
Manhattan	22.1%***	35.3%	28.3%	29.4%
Bronx	19.2%***	15.7%	20.7%	19.8%
Brooklyn	28.5%***	20.6%	26.3%	25.6%
Queens	26.1%***	22.4%	21.0%	21.3%
Staten Island	4.0%***	6.0%	3.8%	3.9%
Current Charges				
Felony	18.7%***	45.1%	27.8%	28.9%
Violent Felony	2.6%***	7.2%	3.9%	4.2%
Domestic Violence	35.2%***	20.0%	27.8%	27.3%
Harm to Property	15.5%***	29.5%	21.9%	22.9%
Harm to Person	56.9%***	30.5%	43.6%	42.4%
Drug *(excl. marijuana)*	8.8%***	19.1%	12.9%	13.5%
Weapon	1.2%*	1.5%	1.5%	1.5%
History of Pretrial Misconduct				
Any Prior FTA	13.5%***	29.9%	19.3%	20.2%
# Prior Cases where People Failed to Appear (mean)	0.29***	0.60	0.43	0.42
Prior Arrests				
Any Arrest	37.2%***	65.5%	49.8%	51.1%
Violation/Infraction	3.1%***	5.7%	4.3%	4.6%
Misdemeanor	32.4%***	55.9%	43.3%	44.2%
Felony	14.7%***	33.1%	20.8%	21.9%
Violent Felony	6.8%***	13.5%	9.4%	10.1%
Drug *(excl. marijuana)*	9.0%***	20.8%	12.8%	13.3%
Harm to Property	12.4%***	28.4%	17.7%	18.7%
Harm to Person	17.0%***	25.3%	22.1%	22.5%
Weapon	2.1%***	4.4%	3.0%	3.1%
Domestic Violence	9.4%***	14.0%	12.3%	12.5%
Sex Charge	0.5%***	0.9%	0.7%	0.8%

	Original Sample		Final Sample	
	Bail Reform Group *(N=16,666)*	**Comparison Group *(N=4,685)***	**Bail Reform Group *(N=12,350)***	**Comparison Group *(N=3,510)***
# Prior Violation/Infraction Arrests (mean)	0.04***	0.07	0.06	0.05
# Prior Misdemeanor Arrests (mean)	0.85***	1.82	1.20	1.25
# Prior Felony Arrests (mean)	0.22***	0.50	0.31	0.33
# Prior Drug Arrests *(excl. marijuana)* (mean)	0.16***	0.39	0.23	0.24
# Prior Harm to Property Arrests (mean)	0.35***	0.84	0.51	0.52
# Prior Harm to Person Arrests (mean)	0.24***	0.38	0.32	0.33
# Prior DV Arrests (mean)	0.13***	0.21	0.18	0.19
Pending Case	10.3%***	23.3%	15.0%	15.9%
Prior Convictions				
Any Conviction	22.6%***	47.9%	31.6%	32.8%
Infraction/Violation	15.6%***	24.5%	21.1%	21.5%
Misdemeanor	11.1%***	32.4%	16.7%	17.9%
Felony	1.8%***	7.3%	2.6%	2.8%
Violent Felony	0.1%***	0.9%	0.2%	0.2%
Drug *(excl. marijuana)*	3.4%***	11.5%	5.2%	5.4%
Harm to Person or Property	7.2%***	21.8%	10.9%	11.7%
Harm to Property	5.7%***	18.1%	8.7%	9.3%
Harm to Person	2.0%***	5.5%	3.1%	3.3%
Weapon	0.4%***	1.4%	0.7%	0.7%
Sex Charge	0.1%**	0.3%	0.1%	0.1%
# Any Prior Convictions (mean)	0.9***	1.97	1.28	1.34
# Prior Violation/Infraction Convictions (mean)	0.04***	0.07	0.06	0.05
# Prior Misdemeanor Convictions (mean)	0.85***	1.82	1.20	1.25
# Prior Harm to Person/Property Convictions (mean)	0.58***	1.21	0.82	0.84
Month of Arraignment				
January	21.5%	22.3%	21.9%	21.9%
February	20.3%***	17.9%	19.1%	19.1%
March	16.8%*	18.4%	17.6%	17.5%
April	11.2%***	15.7%	12.7%	12.7%
May	17.9%***	14.1%	16.4%	16.7%
June	12.3%	11.6%	12.3%	12.2%

*** $p < .001$ ** $p < .01$ * $p < .05$

Exhibit T2: Baseline Characteristics of Bail Reform Group and Comparison Group Samples for the Bail Eligible Pre-Post Analysis, Before and After Inverse Probability of Treatment Weighting (IPTW)

	Original Sample		**Final Sample**	
	Bail Reform Group *(N=3,726)*	**Comparison Group *(N=3,915)***	**Bail Reform Group *(N=3,726)***	**Comparison Group *(N=3,915)***
Demographics				
Age	36.7***	35.4	36.1	36.1
White	12.5%***	8.0%	10.5%	10.0%
Black	48.2%**	51.6%	50.1%	49.8%
Hispanic	32.6%***	36.6%	34.3%	35.0%
Additional Race/Ethnic Groups	6.6%***	3.8%	5.2%	4.9%
Female	18.2%***	7.9%	12.9%	12.6%
Arraignment Borough				
Manhattan	19.0%***	27.4%	23.1%	12.5%
Bronx	15.6%***	18.5%	17.0%	17.3%
Brooklyn	36.3%***	28.7%	32.4%	33.0%
Queens	22.7%**	19.7%	21.4%	20.4%
Staten Island	6.4%	5.7%	6.1%	5.7%
Current Charges				
Felony	74.8%***	91.8%	83.5%*	85.2%
Violent Felony	63.8%***	36.2%	30.9%	29.2%
Domestic Violence	34.9%***	23.7%	29.0%	28.2%
Harm to Property	0.6%**	1.3%	0.8%	1.0%
Harm to Person	52.8%	53.5%	53.0%	54.4%
Drug *(excl. marijuana)*	2.8%***	7.0%	5.2%	5.1%
Weapon	11.2%***	17.9%	14.9%	14.9%
Sex Charge	8.8%***	12.1%	10.0%	10.6%
History of Pretrial Misconduct				
Any Prior FTA	10.6%***	17.8%	13.7%	14.4%
# Prior Cases where People Failed to Appear (mean)	0.19***	0.31	0.25	0.24
Prior Arrests				
Any Arrest	41.2%***	52.2%	46.5%	46.4%
Violation/Infraction	2.6%***	4.7%	3.7%	3.7%
Misdemeanor	36.3%***	44.0%	40.1%	39.7%
Felony	15.2%***	24.6%	19.4%	20.2%
Violent Felony	8.1%***	13.0%	10.7%	10.7%
Drug *(excl. marijuana)*	4.9%***	10.0%	7.4%	7.6%
Harm to Property	9.3%***	13.4%	11.4%	11.6%
Harm to Person	24.0%**	26.8%	24.8%	25.2%
Weapon	2.2%***	4.7%	3.6%	3.4%
Domestic Violence	19.3%	20.1%	18.7%	19.2%
Sex Charge	1.1%	1.3%	1.0%	1.4%

	Original Sample		Final Sample	
	Bail Reform Group *(N=3,726)*	**Comparison Group *(N=3,915)***	**Bail Reform Group *(N=3,726)***	**Comparison Group *(N=3,915)***
# Any Prior Arrests (mean)	1.01***	1.53	1.25	1.28
# Prior Violation/Infraction Arrests (mean)	0.03***	0.06	0.04	0.04
# Prior Misdemeanor Arrests (mean)	0.77***	1.10	0.94	9.92
# Prior Felony Arrests (mean)	0.21***	0.38	0.29	0.29
# Prior Drug Arrests *(excl. marijuana)* (mean)	0.07***	0.15	0.11	0.11
# Prior Harm to Property Arrests (mean)	0.15***	0.29	0.20	0.23
# Prior Harm to Person Arrests (mean)	0.34***	0.41	0.37	0.37
# Prior DV Arrests (mean)	0.31*	0.36	0.32	0.33
Pending Case	13.5%***	17.5%	15.6%	15.7%
Prior Convictions				
Any Conviction	22.5%***	33.8%	28.3%	28.0%
Infraction/Violation	15.8%***	20.6%	18.2%	17.9%
Misdemeanor	9.4%***	19.4%	14.3%	14.7%
Felony	1.2%***	3.7%	2.6%	2.1%
Violent Felony	0.1%***	0.5%	0.3%	0.3%
Drug *(excl. marijuana)*	1.2%***	4.0%	2.5%	2.7%
Harm to Property	2.9%***	6.4%	4.7%	4.6%
Harm to Person	2.6%***	5.5%	4.0%	4.2%
Weapon	0.6%	1.0%	0.8%	0.8%
# Any Prior Convictions (mean)	0.81***	1.20	0.98	1.02
# Prior Violation/Infraction Convictions (mean)	0.03***	0.06	0.04	0.04
# Prior Misdemeanor Convictions (mean)	0.77***	1.10	0.94	0.92
# Prior Harm to Person/Property Convictions (mean)	0.49***	0.70	0.57	0.60
Month of Arraignment				
January	24.6%***	21.2%	22.6%	22.9%
February	20.0%***	16.0%	18.2%	17.3%
March	16.8%	17.1%	17.5%	17.0%
April	11.3%***	15.4%	13.0%	13.6%
May	17.4%	16.1%	17.0%	17.3%
June	9.8%***	14.2%	11.8%	11.9%

*** p < .001 ** p < .01 * p < .05

Exhibit T3: Baseline Characteristics of Bail Reform Group and Comparison Group Samples for the Bail Eligible Contemporaneous Analysis, Before and After Inverse Probability of Treatment Weighting (IPTW)

	Original Sample		Final Sample	
	Bail Reform Group *(N=3,726)*	**Comparison Group** *(N=2,259)*	**Bail Reform Group** *(N=3,456)*	**Comparison Group** *(N=2,121)*
Demographics				
Age	36.7***	35.3	36.0	36.0
White	12.5%***	8.9%	11.0%	10.7%
Black	51.8%***	44.0%	51.6%	52.7%
Hispanic	32.6%	32.0%	32.9%	32.4%
Additional Race/Ethnic Groups	3.0%***	6.6%	4.6%	4.2%
Female	18.2%	6.7%	13.1%	11.8%
Arraignment Borough				
Manhattan	19.0%***	28.4%	23.2%	24.0%
Bronx	15.6%***	19.4%	18.0%	18.4%
Brooklyn	36.3%***	27.5%	33.5%	33.8%
Queens	22.7%***	17.3%	18.2%	16.9%
Staten Island	6.4%	7.4%	7.1%	6.9%
Current Charges				
Felony	74.8%***	94.9%	85.7%***	89.3%
Violent Felony	63.8%***	85.4%	74.9%**	78.1%
Domestic Violence	34.9%***	20.1%	27.9%	26.3%
Harm to Property	0.6%	0.9%	0.7%	0.8%
Harm to Person	52.8%***	61.6%	57.2%	58.8%
Drug *(excl. marijuana)*	2.8%***	1.0%	2.3%	2.4%
Weapon	11.2%***	20.1%	15.9%	16.5%
Sex Charge	8.8%	9.0%	8.6%	8.2%
History of Pretrial Misconduct				
Any Prior FTA	10.6%***	21.5%	14.9%	15.2%
# Prior Cases where People Failed to Appear (mean)	0.19***	0.40	0.27	0.27
Prior Arrests				
Any Arrest	41.2%***	55.2%	45.7%	45.5%
Violation/Infraction	2.6%**	4.0%	3.2%	3.2%
Misdemeanor	36.3%***	45.5%	39.4%	39.3%
Felony	15.2%***	29.3%	20.2%	20.4%
Violent Felony	8.1%***	17.6%	11.4%	11.6%
Drug *(excl. marijuana)*	4.9%***	11.1%	7.3%	7.1%
Harm to Property	9.3%***	16.0%	12.0%	12.1%
Harm to Person	24.0%***	30.8%	26.0%	26.9%
Weapon	2.2%***	4.7%	3.0%	3.3%
Domestic Violence	19.3%	19.1%	18.7%	17.7%
Sex Charge	1.1%*	1.7%	1.2%	1.1%

	Original Sample		Final Sample	
	Bail Reform Group *(N=3,726)*	**Comparison Group *(N=2,259)***	**Bail Reform Group *(N=3,726)***	**Comparison Group *(N=2,259)***
# Any Prior Arrests (mean)	1.01***	1.68	1.24	1.24
# Prior Violation/Infraction Arrests (mean)	0.03***	0.05	0.04	0.04
# Prior Misdemeanor Arrests (mean)	0.77***	1.15	0.90	0.90
# Prior Felony Arrests (mean)	0.21***	0.48	0.30	0.31
# Prior Drug Arrests *(excl. marijuana)* (mean)	0.07***	0.16	0.10	0.10
# Prior Harm to Property Arrests (mean)	0.15***	0.38	0.22	0.24
# Prior Harm to Person Arrests (mean)	0.34***	0.47	0.39	0.37
# Prior DV Arrests (mean)	0.31	0.35	0.31	0.31
Pending Case	13.5%***	19.2%	15.3%	15.1%
Prior Convictions				
Any Conviction	22.5%***	34.5%	26.6%	26.5%
Infraction/Violation	15.8%***	20.5%	17.0%	16.9%
Misdemeanor	9.4%***	19.6%	13.3%	13.1%
Felony	1.2%***	4.2%	2.1%	2.3%
Violent Felony	0.1%**	0.7%	0.3%	0.2%
Drug *(excl. marijuana)*	1.2%***	4.0%	2.4%	2.3%
Harm to Property	2.9%***	7.4%	4.4%	4.4%
Harm to Person	2.6%***	5.4%	3.6%	3.4%
Weapon	0.6%	1.0%	0.7%	0.9%
# Any Prior Convictions (mean)	0.81***	1.25	0.96	0.96
# Prior Violation/Infraction Convictions (mean)	0.03***	0.05	0.04	0.04
# Prior Misdemeanor Convictions (mean)	0.77***	1.15	0.90	0.90
# Prior Harm to Person/Property Convictions (mean)	0.49***	0.84	0.61	0.61

	Original Sample		Final Sample	
	Bail Reform Group *(N=3,726)*	Comparison Group *(N=2,259)*	Bail Reform Group *(N=3,726)*	Comparison Group *(N=2,259)*
Month of Arraignment				
January	24.6%*	21.9%	23.6%	23.8%
February	20.0%	19.7%	19.8%	18.8%
March	16.8%*	14.8%	16.0%	15.8%
April	11.3%	10.4	10.8%	11.0%
May	17.4%	18.9%	18.1%	18.3%
June	9.8%***	14.4%	11.7%	12.2%
Judges				
Judge #1	3.4%***	3.3%	3.6%	3.2%
Judge #2	3.8%***	2.1%	3.3%	3.3%
Judge #3	3.1%	2.3%	2.9%	2.8%
Judge #4	2.7%	2.2%	2.3%	2.0%
Judge #5	3.1%***	1.0%	1.9%	1.5%
Judge #6	2.3%	2.0%	2.3%	2.3%
Judge #7	2.0%	2.1%	2.1%	2.1%
Judge #8	2.3%	1.6%	2.2%	2.4%
Judge #9	2.0%	2.1%	1.9%	1.9%
Judge #10	1.9%	2.1%	2.1%	2.1%
Judge #11	2.3%*	1.3%	1.7%	1.4%
Judge #12	1.8%	1.8%	1.7%	1.6%
Judge #13	1.9%	1.6%	1.5%	1.6%
Judge #14	1.4%*	2.3%	1.6%	1.7%
Judge #15	2.2%***	0.9%	1.7%	1.6%

*** $p < .001$ ** $p < .01$ * $p < .05$

Note: For this analysis, we also balanced the baseline distribution of 15 arraignment judges who each made at least 100 release decisions in the first half of 2020 to reduce bias due to variation in judges' propensities to release individuals without bail.

Section 2: Survival Analysis

We conducted multivariate survival analyses to explore group divergences in re-arrest rates after each successive day out from the initial arraignment date. Specifically, we used Cox proportional hazards regression modeling to analyze the number of days to the first re-arrest, the first felony re-arrest, the first violent felony re-arrest, and the first firearm re-arrest in up to 30 months following people's initial arraignment.

In contrast to logistic regression modeling, which we used to determine *whether* individuals were re-arrested within two years following initial arraignment (we had a minimum two-year follow-up period for every person included in the study), we deployed survival analysis to examine *how soon* people in one group were re-arrested (if at all) compared to people in the other group. Cox regression also allowed us to include in the analysis individuals with different follow-up periods (i.e., the follow-up period ranged from 24 months to 30 months, depending on people's initial arraignment date), as the hazard rates generated by the survival model take into consideration people's time at risk.

Finally, the survival models account for censored data, i.e., observations that are lost to follow-up. In other words, we do not know whether people who had not yet been re-arrested by the end of the follow-up period were re-arrested later. These individuals are "censored" and are counted as having "survived" (not been re-arrested) for the duration of the study. Cox regression modeling uses information from both uncensored and censored data. On the survival plots, censored observations are depicted as vertical lines on each curve. As can be seen on the graphs shown in the main report, censoring starts on day 730, i.e., two years following arraignment, since that was the minimum tracking period we had for each case.

Endnotes

[1] OCA data provided herein does not constitute an official record of the New York State Unified Court System, which does not represent or warrant the accuracy thereof. The opinions, findings, and conclusions expressed in this publication are those of the authors and not those of the New York State Unified Court System, which assumes no liability for its contents or use thereof.

[2] Kuss, O., Blettner, M., & Börgermann, J. (2016). Propensity Score: An Alternative Method of Analyzing Treatment Effects: Part 23 of a Series on Evaluation of Scientific Publications. Deutsches Ärzteblatt International, 113(35-36), 597; Martens, E. P., Pestman, W. R., De Boer, A., Belitser, S. V., & Klungel, O. H. (2008). Systematic Differences in Treatment Effect Estimates Between Propensity Score Methods and Logistic Regression. International Journal of Epidemiology, 37(5), 1142-1147.

[3] Cole, S. R., & Hernán, M. A. (2008). Constructing Inverse Probability Weights for Marginal Structural Models. American Journal of Epidemiology, 168(6), 656-664.

[4] Amoah, J., Stuart, E. A., Cosgrove, S. E., Harris, A. D., Han, J. H., Lautenbach, E., & Tamma, P. D. (2020). Comparing Propensity Score Methods versus Traditional Regression Analysis for the Evaluation of Observational data: A Case Study Evaluating the Treatment of Gram-Negative Bloodstream Infections. Clinical Infectious Diseases, 71(9), e497-e505.

[5] Austin, P. C. (2013). The Performance of Different Propensity Score Methods for Estimating Marginal Hazard Ratios. Statistics in Medicine, 32(16), 2837-2849; Chesnaye, N. C., Stel, V. S., Tripepi, G., Dekker, F. W., Fu, E. L., Zoccali, C., & Jager, K. J. (2022). An Introduction to Inverse Probability of Treatment Weighting in Observational Research. Clinical Kidney Journal, 15(1), 14-20.

[6] Rassen, J. A., Glynn, R. J., Rothman, K. J., Setoguchi, S., & Schneeweiss, S. (2012). Applying Propensity Scores Estimated in a Full Cohort to Adjust for Confounding in Subgroup Analyses. Pharmacoepidemiology and drug safety, 21(7), 697-709.

[7] Chesnaye et al. (2022), Op. Cit.

[8] Beal, S. J., & Kupzyk, K. A. (2014). An Introduction to Propensity Scores: What, When, and How. The Journal of Early Adolescence, 34(1), 66-92.

[9] Stürmer, T., Webster-Clark, M., Lund, J. L., Wyss, R., Ellis, A. R., Lunt, M., ... & Glynn, R. J. (2021). Propensity Score Weighting and Trimming Strategies for Reducing Variance and Bias of Treatment Effect Estimates: A Simulation Study. American Journal of Epidemiology, 190(8), 1659-1670.

[10] Fox, A., & Koppel, S. (2021). Pretrial Release Without Money: New York City, 1987-2020. New York, NY: New York City Criminal Justice Agency (CJA).

[11] Koppel, S., Topel, D., & Bent-Koerick, K. (2021). Annual Report 2019. New York, NY: New York City Criminal Justice Agency (CJA).

[12] Heiler, P. (2020). Efficient Covariate Balancing for the Local Average Treatment Effect. Journal of Business & Economic Statistics, 40(4), 1569-1582.

[13] Chesnaye et al. (2022), Op. Cit.

[14] Degtiar, I., & Rose, S. (2021). A Review of Generalizability and Transportability. arXiv preprint arXiv:2102.11904.

[15] Chesnaye et al. (2022), Op. Cit.

[16] Xiao, Y., Moodie, E. E., & Abrahamowicz, M. (2013). Comparison of Approaches to Weight Truncation for Marginal Structural Cox Models. Epidemiologic Methods, 2(1), 1-20.

[17] Austin, P. C., & Stuart, E. A. (2015). Moving Towards Best Practice When Using Inverse Probability of Treatment Weighting (IPTW) Using the Propensity Score to Estimate Causal Treatment Effects in Observational Studies. Statistics in Medicine, 34(28), 3661-3679.

[18] Stang, A., & Baethge, C. (2018). Imbalance p Values for Baseline Covariates in Randomized Controlled Trials: A Last Resort for the Use of p Values? A Pro and Contra Debate. Clinical Epidemiology, 10, 531.

[19] Elze, M. C., Gregson, J., Baber, U., Williamson, E., Sartori, S., Mehran, R., ... & Pocock, S. J. (2017). Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. Journal of the American College of Cardiology, 69(3), 345-357; Campbell, C. M., & Labrecque, R. M. (2022). Panacea or Poison: Assessing How Well Basic Propensity Score Modeling Can Replicate Results from Randomized Controlled Trials in Criminal Justice Research. Journal of Experimental Criminology, 1-25.

[20] Rassen et al. (2012), Op. Cit.

DATA COLLABORATIVE FOR JUSTICE

AT JOHN JAY COLLEGE

STRENGTH IN NUMBERS

JOHN JAY

COLLEGE OF CRIMINAL JUSTICE