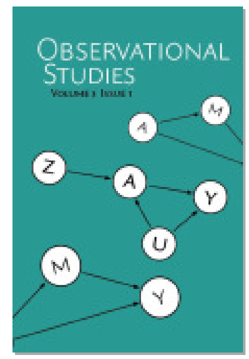
The causal impact of bail on case outcomes for indigent defendants in New York City

Kristian Lum, Erwin Ma, Mike Baiocchi

Observational Studies, Volume 3, issue 1, 2017, pp. 38-64 (Article)

Published by University of Pennsylvania Press
DOI: https://doi.org/10.1353/obs.2017.0007



Observational Studies 3 (2017) 39-64 Submitted 5/17; Published 10/17

The causal impact of bail on case outcomes for indigent defendants in New York City

Kristian Lum kl@hrdag.org
Human Rights Data Analysis Group
109 Bartlett St. # 204
San Francisco, CA 94110, USA

Erwin Ma EZMa@legal-aid.org
New York Legal Aid Society
199 Water St.
New York, NY 10038, USA

Mike Baiocchi baiocchi@stanford.edu
Stanford Prevention Research Center
1265 Welch Rd
Palo Alto, CA 94305, USA

Abstract

It has long been observed that defendants who are subject to pre-trial detention are more likely to be convicted than those who are free while they await trial. However, until recently, much of the literature in this area was only correlative and not causal. Using an instrumental variable that represents judge severity, we apply near-far matching– a statistical methodology designed to assess causal relationships using observational data– to a dataset of criminal cases that were handled by the New York Legal Aid Society in 2015. We find a strong causal relationship between bail– an obstacle that prevents many from pre-trial release– and case outcome. Specifically, we find setting bail results in a 34% increase in the likelihood of conviction for the cases in our analysis. To our knowledge, this marks the first time matching methodology from the observational studies tradition has been applied to understand the relationship between money bail and the likelihood of conviction.

KEY WORDS: pre-trial detention, money bail, bail, observational studies, causal inference, near-far matching, matching

1. Introduction

In the United States, the money bail system has come under recent scrutiny due to its contribution to mass incarceration and its impact on poor defendants (Human Rights Watch, 2017). Under the current money bail system in New York City, a judge may choose to set an amount of money (bail) that an individual is required to pay in order to secure his release from detention prior to standing trial. If the defendant can pay the bail– either from personal or familial funds or by hiring a bail bondsman, which requires paying both a non-refundable fee and posting sufficient collateral – the defendant is released from detention. If the accused cannot raise the funds, he must remain incarcerated until the case is resolved, whether by plea or trial. At the most basic level, the money bail system disproportionately impacts the poor because they are generally less able to afford either the bail amount in full or to hire a bail bondsman, thus subjecting them to different pre-trial conditions than those with greater financial resources.

There are several ways in which the effects of these differential conditions extend beyond the pre-trial stage of the criminal justice process. For example, several studies have found that even short periods of pre-trial detention are associated with substantially higher rates of recidivism (Lowenkamp et al., 2013; Gupta et al., 2016; Phillips, 2007a). Though Dobbie et al. (2016) did not find a significant relationship between pre-trial detention and recidivism, they did find that pre-trial detention reduced earnings and rates of employment. It has also been suggested that pre-trial detention causes a higher likelihood of conviction based on the observation that those who are detained pre-trial are more likely to plead or be found guilty than those who were released (in some cases even after controlling for many legally relevant factors (Ares et al., 1963; Rankin, 1964; Phillips, 2007b, 2008)).[1] For example, a 1972 report demonstrated systematic differences in case disposition between those who had been released and those who had been detained prior to trial (Warren et al., 1972). Despite not using formal causal inference methods, the authors argued the unconstitutionality of bail practices on the grounds that, because the difference between pre-trial detention or release was often a matter of the financial means of the defendant and this distinction has causal impact on the outcome of the trial, the administration of bail unfairly discriminated against poor defendants. This report was used as supporting evidence for bail reform measures in the 1970s.

More recent work has applied modern methods for causal inference to assess the role played by the money bail system in the ultimate disposition of the case. Gupta et al. (2016) employ a measure of judge strictness as an instrumental variable to estimate the causal impact of setting bail on case outcome to data from Philadelphia and Pittsburgh. They find that assigning money bail increases the likelihood that the defendant is found guilty by 12%. Leslie and Pope (2016) applies a similar two-stage instrumental variable method to national-level data and data from New York City to assess the causal impact of pre-trial detention on case disposition. Leslie and Pope (2016) also uncovered a "strong causal relationship." Stevenson (2016) and Dobbie et al. (2016) employ similar methodology to data from Pennsylvania and Miami-Dade County. Both find a statistically significant impact of the money bail system on the outcome of the case.

In this paper, we also bring modern statistical methods to bear on the problem of quantifying the impact of setting bail on case outcome. Similar to Warren et al. (1972), our analysis focuses on the population of clients represented by the New York Legal Aid Society (NYLAS)– one of the groups of public defenders in New York City. We believe the population of defendants represented by the NYLAS is an important subgroup on which to focus because those who cannot afford their own attorney are also at risk of not being able to post bail if it is set, making this a particularly vulnerable population deserving of specific

1. For an excellent overview of studies regarding the relationship between pre-trial detention and case outcome (including references to some studies that found no statistically significant relationship), see Phillips (2007b).

attention. Thus, our focus on indigent defendants differentiates our work from similar recent studies, such as Gupta et al. (2016); Leslie and Pope (2016); Dobbie et al. (2016); Stevenson (2016), which focus on the effect of the money bail system on the population at large, regardless of their representation.

The work we present is similar to the mentioned studies in that we also use an instrumental variable that is a measure of judge severity. However, as opposed to the two-stage model-based methods used by the other studies, we use near-far matching (Baiocchi et al., 2012; Lorch et al., 2012)– a procedure to pair nearly identical defendants who are facing identical charges in the same county. The paired defendants differ only in that they were arraigned by judges with varying levels of severity. This procedure ensures balance and common support in the covariate space between the comparison groups, forces close examination of the IV, and "strengthens" the instrumental variable leading to more reliable inference. To our knowledge, this is the first use of matching techniques in this setting.

2. The criminal justice process in New York City

To begin, it is helpful to understand how arrested individuals are processed in New York City. Upon arrest the suspect is taken to a police precinct. People who were arrested for very minor, non-violent crimes are often released with a desk appearance ticket that specifies a later date at which they are to appear in court. Those who are not released with a desk appearance ticket are brought to central booking to await arraignment. There, the New York Criminal Justice Agency (CJA) administers an interview in which basic demographic information, criminal history, and other personal details are recorded. This information is then fed into a formula that outputs a "risk score", which is meant to predict whether the individual will appear in court if released. Then, they are assigned a defense attorney who will represent them at arraignment.

Under New York law, the arraignment is to take place within twenty-four hours of booking. At the arraignment, the district attorney presents the charges against the defendant, briefly gives a summary of the evidence, and makes a recommendation to the judge regarding whether bail should be set. The defense attorney then argues why the defendant should be released on his own recognizance before his trial or that bail should be set at a reasonable amount. Based on this information, combined with the CJA's risk score, the judge then makes a decision to dismiss the case, release the defendant on their own recognizance, set bail, or in some serious cases remand the defendant directly to custody without offering bail. For our purposes, we consider the remanded cases as having been assigned infinite bail. In our analysis we have two treatment levels - "bail set" and "bail not set," so in our analysis remanded defendants are considered as having bail set.

It is important to note that an arraignment is not a trial. The majority of what is said is perfunctory. The prosecution has an enormous advantage in that they've been able to interview police and other witnesses whereas the defendant just met his lawyer for the first time. On average, based on the experience of attorneys who have worked in these courts for many years, we estimate that a typical arraignment lasts between three and ten minutes.

Another detail that will be central to our study design is that defendants are essentially randomly assigned to judges. Defendants appear before whichever judge happens to be on duty at the time of their arraignment– anywhere from zero to a bit over twenty four hours after their arrest. During busy times, the time from arrest to arraignment might be on the upper end of that range, and during times when fewer arrests have been made, the arraignment likely takes place more quickly. This range encompasses several judges' shifts, so which judge hears the arraignment will depend both on the time of arrest as well as the how busy the court is on that day, the latter effect being completely external to the relevant facts of each particular case. Other factors that affect the time to arraignment are whether the defendant needed medical care, whether and how long they were interrogated, and the quickness with which the Department of Corrections staff is working on that particular day. Thus, even if it were the case that arrests that take place at night are systematically different from those that take place during the day, the defendants will be pseudo-randomized to judges based on, among other factors, the volume of cases being processed on that day. Additionally, the processes by which judges are assigned to arraignments and trials are completely separate. Which judge oversees the arraignment has no bearing on which judge will preside over the trial, should a trial take place (an unlikely event), or any other aspect of the process that leads to a final disposition. That is, the arraignment judge has no involvement in plea bargaining, enforcing discovery deadlines, deciding legal motions, or other factors that can influence a case's disposition. Thus, we believe it is reasonable to assume that only effect an arraigning judge is certain to have on a case outcome is her bail decision at arraignment.

After the arraignment, those defendants who are released on their own recognizance are free to go. Those for whom bail was set are held at the court for a few hours to see if a friend or relative will show up to post bail for them. The friend or relative may also contract with a bail bondsman to post the bond amount, which would require the posting of collateral and the payment of a fee which is limited to 10% by statute. That fee is non-refundable, even if all charges against the defendant are eventually dismissed. After a few hours, those defendants who are not able to post bail are taken to Rikers Island until their case is resolved, either by dismissal, a plea deal, or a trial.

Once at Rikers Island, the defendant has many incentives to plead guilty that do not apply to those who await trial while free. According to a recent report (Human Rights Watch, 2017), detained defendants often plead guilty to secure their immediate release, accepting convictions for which the punishment is "time served". According to numerous defendants interviewed by Human Rights Watch, immediate release is desirable (in some cases, even at the expense of pleading guilty to a crime they did not commit) because the financial costs of remaining imprisoned are too high. The report details cases in which detained defendants have lost their jobs and subsequently their housing and vehicles while awaiting trial. One defendant's son was even taken to foster care because the father could not find alternative child care while he was imprisoned awaiting trial. The financial burden of imprisonment is borne not only by the imprisoned but also their family members, who often must continue to pay their bills without the contribution of the imprisoned defendant. This social pressure may also play a role in persuading the defendant to plead guilty when the facts of the case would not have led to incarceration.

Additionally, some may plead guilty simply to escape the harsh conditions at Rikers Island. Bharara (2014), a report from the U.S. Department of Justice, detailed the "culture of violence" that is pervasive at Rikers Island, citing "a pattern and practice of using unnecessary and excessive force against inmates," and inmate-on-inmate violence that results from lack of appropriate supervision. Last, of the few cases that got to trial, incarcerated clients are always seated next to their attorneys and directly behind them are two uniformed court officers who sit there at all times, potentially giving them the appearance of a criminal and influencing how they were perceived by jurors. Jurors for trials of released clients face no such influence, as the released defendants not seated with uniformed officers.

Taken in total, there is ample contextual information suggesting several plausible mechanisms by which pre-trial detention could *cause* a defendant to be more likely to plead or be found guilty.

3. Data

i. Inclusion/exclusion

Our dataset consists of all felony and misdemeanor cases that began in 2015 and were handled exclusively by an attorney from NYLAS. We do not consider cases in which the defendant was extradited, the case was transferred to a special court (e.g. family court), very irregular cases (e.g. the crime was abated by the death of the client), or cases where non-NYLAS counsel was assigned. We also do not consider cases that were disposed at arraignment, i.e. cases in which the defendant immediately plead guilty or the judge dismissed the charges at arraignment. We do not consider these cases to be part of the population of interest because in these cases bail cannot be set and the concept of pre-trial detention is irrelevant, as there is to be no further trial and there is no pre-trial period of which to speak. In our dataset, we had 86,912 cases in 2015 that fit these criteria. After omitting some cases that contained obvious recording errors, those for which the county in which the case was adjudicated was missing, and duplicate entries, we are left with 85,329 cases for analysis.

In the analysis presented in the main text below, we include cases that received an "adjournment in contemplation of dismissal" (ACD) at arraignment. For these cases, bail is not set and the defendant is released. The case will automatically be dismissed under the condition that the defendant has no more trouble with the law for the duration of the specified period, typically six months or one year. ACDs are sometimes granted for low-level offenses by first-time offenders, and the vast majority result in dismissals. Although these are essentially deferred dismissals at arraignment, some defendants are not able to meet the requirements of the ACD and they are eventually found guilty of the original charges. Given the ambiguity of whether these should be considered part of the population of study, in the appendix we repeat the analysis presented below excluding all of the ACD cases. To be clear, even when the ACDs are included in the analysis, they are counted as cases in which the defendant was released, not detained, pre-trial. In general, the findings from excluding the ACD cases are substantively similar, though the reduced sample size that results from dropping the ACD cases degrades statistical power. Because of this, in this supplementary analysis, some results that were statistically significant when including ACDs are no longer so.

ii. Covariates

Our dataset includes a variety of demographic covariates about the individual – age at the time of the alleged crime, gender, race, and ethnicity. At the time of processing in the intake interview, the defendant is also asked to report his employer, weekly income, phone number, and address. We include an indicator of whether the defendant declared an employer, the value of the self-reported weekly income, and indicators of whether a phone number and address was reported. Although this is all self-reported information, this is the same information that is available at the time of arraignment, and thus is the information relevant to determining whether a bail will be set. Last, as a measure of prior criminal activity, we include the number of prior counts for which NYLAS represented that client in the previous year, 2014 (denoted by 'Records 2014'). This is a noisy measure of prior criminal activity, as it is possible that in 2014, the defendant had additional charges but had different legal representation. It is important to note that despite the fact that there may be relevant variables (like a longer criminal history variable) that are omitted from the analysis, the instrumental variable method we employ allows us to obtain causal estimates nonetheless. This is analogous to a medical study in which patients are paired along all variables that are measured and then randomized to the treatment or control group to account for all remaining unmeasured variables.

For each case, our dataset also includes information about the charges against the defendant– the type of the offense (misdemeanor or felony), the most serious charge against the defendant (the "top charge"), the class of the most serious charge (A, B, C, D, E). The top charge in the case is a specific category that describes both the nature of the crime as well as the severity, in most cases denoted by the number following the crime description, e.g. "Assault 3". The top 20 most common charges are given in Table 1, along with the number of cases for which each was the top charge in our dataset.

One last variable that requires additional explanation is the county in which the charges were booked. In New York City, counties map one-to-one to boroughs (with which we expect most readers are more familiar). Specifically, New York County is co-extensive with Manhattan, Bronx County with the Bronx, Kings County with Brooklyn, Queens County with Queens, and Richmond County with Staten Island. In general, we refer to counties (except when notationally inconvenient), though for our purposes, counties should be considered interchangeable with boroughs.

iii. Outcome and treatment

Outcome

The outcome variable of interest, G, is an indicator of whether the defendant was found guilty. Specifically, we set $G = 1$ if the final disposition of the case was a guilty plea (the defendant plead guilty without going to trial) or a guilty verdict (the case went to trial, and the defendant was found guilty), and $G = 0$ if the case was dismissed (the charges were dismissed without going to trial) or the defendant was acquitted (the case went to trial, and the defendant was found not guilty). We treat G as missing if a final determination has yet to be made.

Contrary to the common Hollywood depiction of criminal justice proceedings, cases very rarely go to trial– of the approximately 80,000 cases in our dataset, fewer than 200 of them

	Top Charge	Count
1	ASSAULT 3	12047
2	THEFT OF SERVICES	9515
3	CRIM POSSESSION OF STOLEN PROPERTY 5	7017
4	CRIM POSSESS OF MARIJUANA 5	4396
5	CRIM POSSESS CONTROLLED SUBSTANCE 7	3960
6	CRIM POSSESSION OF A WEAPON 4	3732
7	AGGRAVATED UNLICENSED OPERATION 3	3408
8	CRIM MISCHIEF 4	2484
9	CRIM CONTEMPT 2	2027
10	ASSAULT 2	1691
11	CRIM OBSTRUCT BREATH BLOOD	1484
12	CRIM POSSESS CONTROLLED SUBSTANCE 3	1249
13	CRIM TRESPASS 3	1233
14	AGGRAVATED HARASSMENT 2	1217
15	RESISTING ARREST	1210
16	OPERATE MV - INTOXICATED	1161
17	CRIM TRESPASS 2	1149
18	CRIM SALE OF MARIJUANA 4	1066
19	CRIM SALE CONTROLLED SUBSTANCE 3	1023
20	CRIM POSS FORGED INSTRUMENT 2	1011

Table 1: Twenty most common top charges.

went to trial. This underscores the importance of desicisions made at arraignment. A histogram of the time until trial for those cases that have gone to trial is shown in Figure 1.

Treatment

As the treatment variable, T, we use an indicator that denotes whether bail was set in the case. Although our hypothesis for how the money bail system causes worse case outcomes revolves around pre-trial detention– not the setting of bail– pre-trial detention only occurs if bail is set and the defendant cannot pay. Thus the proximate cause of pre-trial detention is the setting of bail. Also, we believe that the setting of bail is the appropriate treatment variable in this case, as that is the action taken by the judge and the point at which intervention can be taken. Our argument in this regard is similar to that presented in Gupta et al. (2016).

Relationship between outcome and treatment variables

Table 2 shows the proportion of cases that resulted in a conviction conditional on whether bail was set, i.e. $Pr(G = 1 \mid T)$. In the next section, we provide deeper detail on how we account for missingness in our analysis. However, for the descriptive analysis provided in Table 2, we simply exclude cases that have not yet concluded. From this table, it is clear that cases for which bail was set were much more likely– in some cases more than twice as

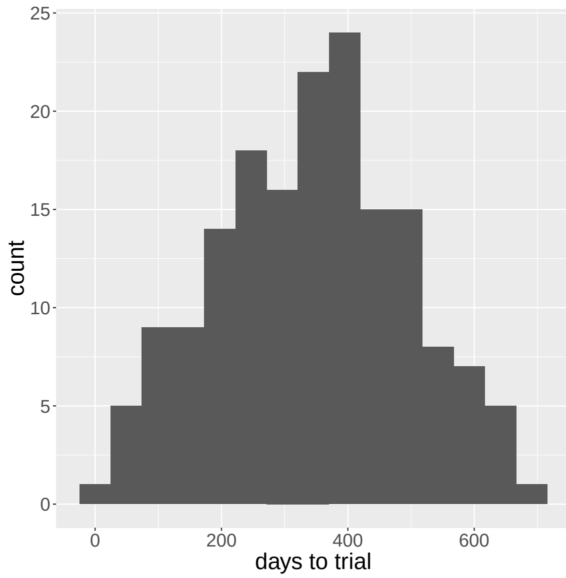


Figure 1: Histogram of the number of days from arrest to trial *for cases that have already gone to trial*.

likely– to result in a determination of guilt than those for which bail was not set. If bail had been entirely randomly assigned, as in a randomized trial, then Table 2 would provide extraordinarily strong evidence that setting bail causes an increase in the probability of a determination of guilt. Instead of randomly setting bail, the process by which bail is assigned uses information about the facts of the case, and defendant information, and this could impose a non-causal connection between bail setting and the determination of guilt.

The relationship between case-specific information and bail setting is illustrated in Table 3, which shows standardized differences in the covariates by whether or not bail was set. In this table, the covariates are separated from the outcome, treatment, and IV by a horizontal line. From this, it is clear that the population for which bail is set is systematically different than that which is released without bail. In particular, the defendants that are assigned bail are, on average, older, more likely to be black, male, and (unsurprisingly) have a criminal history than those who are not.

iv. Instrumental variable

For the instrumental variable (IV), we calculate a measure of judge strictness or severity. To conform to the conventions of near-far matching, we calculate this so that low levels of the IV correspond to more strict judges, and higher levels of the IV to more lenient judges. We note that by "strict" and "lenient", we do not mean to imply that stricter judges adhere more strictly to the law and the lenient judges are deviating from the legally prescribed course of action. We use this terminology simply to differentiate those judges who are more likely to set bail and those who are less likely.

Stratification	Stratum	No Bail	Bail
aggregate	total	0.33	0.73
County	Bronx	0.33	0.63
	Kings	0.28	0.68
	New York	0.33	0.77
	Queens	0.37	0.80
	Richmond	0.45	0.71
Crime Type	Felony	0.57	0.74
	Misdemeanor	0.30	0.72
Gender	Female	0.25	0.74
	Male	0.35	0.73
	NULL	0.18	0.78
Age Bin	[0,18)	0.17	0.59
	[18,25)	0.25	0.67
	[25,35)	0.35	0.69
	[35,45)	0.38	0.75
	[45,55)	0.42	0.80
	[55,65)	0.40	0.80
	65+	0.34	0.75
Any Priors	no prior record	0.30	0.69
	prior record	0.43	0.76

Table 2: Proportion found guilty (by plea or conviction) by whether bail was set at arraignment.

There is ample evidence to support the idea that judges make rulings not only based on the facts of the case, but also based on personal factors, i.e. time since the judge's last meal break or whether their favorite sports team has recently experienced an unexpected loss (Danziger et al., 2011; Eren and Mocan, 2016).[2] Although these studies only address whether individual judges become more or less lenient based on external factors, these studies speak to the fact that the decisions are not exclusively functions of the facts of the case. Given that, it is reasonable to believe that there are idiosyncratic, dispositional differences in judges which cause some to be more lenient and some to be more strict.

Several other analyses have used judge severity as a pseudo-randomizer (Martin et al., 1993; Aizer and Doyle, 2015; Kling, 2006). In particular, Kleinberg et al. (2017) use a similar instrumental variable in an analysis how judges in New York City determine to whom they grant pre-trial release and the likelihood with which individuals who would have been released would have failed to appear in court. Other studies also rely on judge randomization or quasi-randomization in assessing the causal impact of incarceration or probation on recidivism (Green and Winik, 2010; Berube and Green, 2007). At the core of all of these analyses is the assumption that some part of the decision-making process

2. Weinshall-Margel and Shapard (2011) gives a compelling critique of Danziger et al. (2011) and shows that the temporal pattern in decisions may be the result of intentional ordering of the cases.

	No Bail n=62826	Bail n=14057	Abs St Dif
Guilty	0.33	0.73	1.04
Bail Set	0.00	1.00	2.04
IV	0.01	-0.02	0.22
Age	31.93	35.94	0.32
White	0.29	0.28	0.02
Black	0.48	0.61	0.26
Non-Hispanic	0.63	0.66	0.08
Male	0.79	0.90	0.30
Prior Records 2014	0.38	1.36	0.70
Weekly Income	67.41	63.94	0.02
Any Income	0.15	0.14	0.02
Reported Employer	0.21	0.21	0.00
Reported Phone Number	0.18	0.19	0.04
Reported Address	0.89	0.93	0.12

Table 3: Table of pre-match standardized differences.

depends on features of the judge, rather than the facts of the case, such that two judges when presented with identical cases could reach different conclusions about whether or not bail should be set. Our identification strategy makes use of the insight that some judges are predisposed to set bail ("strict") and others are less likely ("lenient"). More technically, our "pseudo-randomizer" is a judge's rate of granting pre-trial release without bail, for a specific crime type, relative to other judges in that borough. The "exclusion restriction" is the assumption that, as we have previously argued, the judge affects the outcome of the case only via her decision to set or not set bail at arraignment. We return to this assumption in the discussion.

We calculate the instrumental variable separately for each borough and top charge, resulting in a judge-borough-crime-specific measure of severity. Similar to Gupta et al. (2016); Leslie and Pope (2016), we use a leave-one-out method for calculating this variable so that the ith defendant's own outcome does not influence the calculation of the instrumental variable for his case. Let T_{jbci} denote the treatment variable (1 if bail was set; no otherwise) of the ith individual seen by judge j in borough b with top charge c. We define $T'_{jbci} = 1 - T_{jbci}$. Then, we calculate judge severity measure for the i^*th defendant as follows:

$$S_{jbc}^{(i*)} = \frac{1}{n_{jbc} - 1}\left(\sum_i T'_{jbci} - T'_{jbci*}\right) - \frac{1}{n_{bc} - 1}\left(\sum_{i,j} T'_{jbci} - T'_{jbci*}\right), \tag{1}$$

where n_{jbc} is the number of cases with top charge c seen by judge j in borough b, and n_{bc} is the number of cases with top charge c seen in borough b. The first half of this sum is a leave-one-out measure of the rate at which the jth judge releases defendants without bail for cases of type t in borough b. The second term re-centers the instrumental variable such that it is interpretable as the judge's deviation in severity for that crime relative to the average for that crime in the same borough. We omitted cases for which the judge had seen only one of its type, as this would make calculating the IV impossible under the leave-one-out calculation. This results in a sample size of n =80,084 after the omission.

Table 4 shows the mean of each of the covariates, grouped by the quartile into which the instrumental variable associated with that case fell. Continuous variables are accompanied by a standard deviation in parenthesis; binary variables are not. This table is provided to demonstrate that the IV is substantively unrelated to the measured covariates, and in doing so, support the assumption that the IV is also uncorrelated with additional confounding variables. Of course, this assumption cannot be checked empirically, and we must rely upon understanding of the process by which arraignments are distributed among judges to assert that the IV can serve as a reasonable pseudo-randomizer. In this table, the magnitude of differences between quartile groups should be interpreted in the context of their natural variability, i.e. the standard deviation for continuous variables. Our IV does not vary substantially by IV along any of the covariates, indicating that the IV is a reasonable randomizer. However, Black and Male exhibit the largest discrepancies by group relative to the others, so we pay special attention after matching to those variables to be sure that they are well balanced.

	mean (sd)	Q.1	Q.2	Q.3	Q.4
Age	32.61 (12.53)	33.74	31.32	31.75	33.59
Black	0.5	0.53	0.51	0.49	0.49
White	0.29	0.29	0.27	0.28	0.3
Non-Hispanic	0.63	0.65	0.63	0.62	0.63
Male	0.81	0.81	0.83	0.8	0.78
Reported Employer	0.21	0.24	0.22	0.2	0.2
Reported Phone Number	0.18	0.2	0.17	0.17	0.19
Reported Address	0.9	0.9	0.9	0.89	0.89
Weekly Income	66.32 (222.63)	78.27	64.04	61.27	61.55
Any Income	0.15	0.16	0.15	0.14	0.14
Prior Records 2014	0.56 (1.39)	0.68	0.48	0.47	0.59

Table 4: Summary of Covariates by IV Quartile

4. Method

Current recommendations for best practices in observational studies of medical interventions typically favor a matching approach, rather than the two-stage model-based inference that is popular in econometrics (Hickam et al., 2013). In this study, we use near-far matching– very roughly speaking, a combination of propensity score matching with instrumental variables. Near-far matching differs from propensity score matching in two primary ways. First, near-far matching does not match between exposed and unexposed (a.k.a. treated and control or bail set and no bail set); it matches between "encouraged" and "unencouraged." Second, in propensity score study designs, the analyst makes the assumption that the observed covariates are sufficient for unbiased estimation of selection into exposure (this assumption goes by several names: strongly ignorable treatment assignment, selection on the observed, and exchangeability). Near-far matching bypasses this assumption and instead incorporates an instrumental variable that accounts for pseudo-randomization of subjects.

The logic of near-far matching follows the design of a randomized experiment that suffers from noncompliance with the randomization - this is sometimes called an encouragement design (Holland, 1988). In encouragement randomized trials, some physicians are randomly assigned to be encouraged to perform or suggest a particular treatment to their patients, others are not (Dexter et al., 1998). The result is that some patients, even after accounting for their own personal attributes or the severity of their condition, are more likely to receive the treatment due only to the level of encouragement their physician received. Analogous observational studies in which there is patient-independent variability in the physician-specific inclination towards a treatment can be undertaken in this setting if patients are pseudo-randomized to physicians. Methodology for these studies exploits this randomized push towards receiving the treatment to isolate the natural experiment that exists in the data (Zubizarreta et al., 2014).

The machinery developed for observational analogues to encouragement randomized trials, such as near-far matching, is applicable to our data as well. In the canonical case, whether a patient sees an encouraged or unencouraged physician is random because physicians are randomly assigned to encouragement groups. By analogy, we have argued that whether a defendant sees an "encouraged" (strict) judge or an "unencouraged" (lenient) judge is random due to the process by which arraignments are scheduled. Although judges themselves aren't encouraged to be stricter or more lenient by researchers performing an experiment, the analogy still holds. Because of personal, dispositional characteristics of the judge that are independent of the attributes of the defendant and their case, identical defendants who see stricter judges are more likely to have bail set than those who see more lenient judges.

Near-far matching mimics a randomized encouragement trial by preferentially creating matched pairs of observations that are (i) as nearly identical in pre-exposure variables as possible ("near in covariates"), while (ii) being as dissimilar as possible in their pseudo-randomized push to either be exposed or unexposed ("far in their encouragement"). Pairwise covariate proximity is measured by calculating the Mahalanobis distance between covariate vectors. A non-bipartite matching algorithm is then used to find a set of pairings that minimizes the Mahalanobis distance between the matched pairs while maximizing the pairwise difference in the instrumental variable. In our case, this would look like finding two identical defendants – that is, who looked the same in all ways measured in our data set prior to the bail-setting hearing, but one defendant was routed to a strict judge and the other defendant was routed to a lenient judge. Note that within this pair we are attempting to isolate the judge's predisposition and use it as the determining factor for bail-setting, rather than allowing differences in the facts-of-the-case being the determining factor.

Matching-based study designs focus heavily on the task of identifying reasonable comparator groups and limiting the analysis to those observational units. That is, we exclude observations because the real world data set did not give rise to suitable comparators - much like in a medical study when certain people are excluded from (or always given) a particular treatment (Lu et al., 2001). In our case, you might imagine that there are very few 80 year old women in Queens who are accused of Assault 3 whose arraignment was done by a strict judge. These defendants would not have reasonable contrasts (80 year old women in Queens who are accused of Assault 3 whose arraignment is done by a lenient judge) and thus cannot inform our causal discussion about the impact of having bail set because there is no information about the outcomes of these types of defendants when they are seen by more lenient judges and are thus more likely to be released on their own recognizance. This process of finding suitable comparators and eliminating unreasonable or uninformative units strengthens our instrumental variable, thus preventing unreliable inferences derived from weak instrumental variables. For an in-depth discussion of the problems associated with using weak instrumental variables, see Bound et al. (1995); Imbens and Rosenbaum (2005).

Most modern matching algorithms have a optimal ways for finding the most dissimilar or uninformative units and removing them from the analysis. In our study we use sinks– "phantom" observational units that have the unique property that they are perfect matches to all real data points. The matching algorithm then runs on the augmented data set– real and phantom observations. The algorithm will tend to pair hard-to-match observational units to the sinks. In this implementation of near-far matching, we automatically select the optimal number of sinks by maximizing the F-statistic of a hypothesis test that measures the strength of the instrumental variable, i.e. the degree to which encouragement correlates with treatment assignment. If a real observation is matched to a sink then we remove that observational unit from our analysis.

The output of the matching procedure is a set of matched pairs, $\{i_1, i_2\}$ for $i = 1, ..., I$, where i_1 and i_2 are the indices of the encouraged and unencouraged defendants, respectively, in the ith pair. Then, for example, G_{i_1} and G_{i_2} are the case outcomes for the ith matched pair. Similarly, T_{i_1} and T_{i_2} are the the indicators of whether bail was set for the encouraged and unencouraged defendants, respectively, in the ith matched pair. Although not immediately obvious, the instrumental variable is embedded in the subscript notation, as those defendants who had high values of the IV are assigned to i_1 and those with low levels of the IV to i_2. Having obtained matched pairs, inference is then relatively straightforward.

The causal relationship is measured by estimating the effect ratio as,

$$\lambda = \frac{\sum_{i=1}^{I} G_{i_1} - G_{i_2}}{\sum_{i=1}^{I} T_{i_1} - T_{i_2}}.$$

This estimator is very similar to the Wald estimator commonly used in the econometric and social science literature. Unlike the estimator derived from a two-stage least squares approach, λ does not incorporate IV-derived weights into the estimate.[3] In technical terms, one can describe λ as a complier average causal effect of the risk difference which is conditional on the matched set. It measures the ratio of the difference in outcome between the encouraged and unencouraged groups to the difference in treatment. More informally, one can describe the estimate as the increase in probability of conviction *due to bail setting* for those defendants whose bail determination was likely to switch based on the type of judge that presided over the arraignment. That is, for the subset of cases for which judges

3. In the two-stage least squares setting, the standard estimator weights each observation in proportion to the magnitude of its IV. In contrast, the estimator here does not use information about the relative difference between the IVs in each matched pair. Weights could be applied, and there are good arguments for doing so. However, we believe making the IV binary within matched-pair more closely follows the randomized control trial framework, where noncompliance depends on encouragement to follow assignment. For more discussion of this choice of estimator in this setting, see Baiocchi et al. (2010).

might come to opposite determinations, λ measures the causal effect of bail on likelihood of conviction. For example, for crimes for which all defendants for that crime were assigned bail, it is not possible to assess the effect of setting bail because we have no instances of defendants who were accused of that crime who were released on their own recognizance to which to compare. Thus, our inferences do not pertain to crimes that are so minor that no judge would ever set bail nor to crimes that are so heinous that all judges would necessarily set bail. Confidence intervals are calculated using a permutation test, followed up with a sensitivity analysis that assesses the level of residual bias that is required to nullify the conclusions. For details, see Baiocchi et al. (2010).

The specific details of our matching procedure are as follows. We first stratify all defendants in our dataset into top charge-county-gender groups. These are the groups within which pairs will be created, forcing an "exact match" on the top charge, county, and defendant gender of the case. For example, female defendants in New York County whose top charge was Criminal Mischief 2 will only be matched to other female defendants in New York County whose top charge was Criminal Mischief 2. Within these groups, we use the nearfar package in the R computing environment to match similar defendants to one another (Rigdon et al., 2016). The output of this procedure is a set of pairs of same-gender defendants who are each accused of identical crimes in the same county and who are maximally similar on all other covariates. Paired defendants differ in that they were arraigned by judges with differing levels of severity. Not all defendants are paired– some are dropped as described to achieve the best possible inference.

5. Results

Covariate balance

After the matching procedure is complete, we are left with n = 61,486 defendants in our study. (See Figure 3.) We first assess whether our matching algorithm has successfully achieved covariate balance between the two groups. This is shown in Table 5, which show the standardized differences for each covariate. We find that we were able to obtain excellent balance, including on the Black and Male variables which have absolute standardized differences of 0.00. For all covariates (i.e. all variables except the treatment, IV, and outcome), we attained a standardized difference of less than 0.01. That is, the average difference between the values of each variable for paired defendants was less than one percent of one standard deviation for that variable. This far exceeds the accepted standard that the standardized differences ought to be less than 0.10 (Silber et al., 2001). These tables do not include a charge or county variable because defendants were matched only to other defendants who shared the same top charge and county. So, in some sense, these tables understate the degree of balance by not explicitly showing that we have attained perfect balance on top charge and county.

A before and after covariate balance comparison is shown in Figure 2. This plot is inspired by the commonly used Love plots, which show the standardized differences between treatment and control groups pre- and post-matching. We interpret the core function of the Love plot to be a display of the improvement in covariate balance when moving from a naïve analysis (i.e., all treated vs all control) to a thoughtfully designed study. In this spirit, we display a comparison of covariate balance in the naïve, unmatched setting (defendants

	Strict n=28367	Lenient n=28367	Abs St Dif
Guilty	0.41	0.40	0.03
Bail Set	0.21	0.16	0.12
IV	-0.07	0.07	1.18
Age	32.69	32.71	0.00
White	0.28	0.28	0.00
Black	0.52	0.52	0.00
Non-Hispanic	0.65	0.65	0.00
Male	0.81	0.81	0.00
Prior Records 2014	0.54	0.53	0.00
Weekly Income	53.00	52.75	0.00
Any Income	0.12	0.12	0.00
Reported Employer	0.17	0.17	0.00
Reported Phone Number	0.15	0.15	0.00
Reported Address	0.91	0.91	0.00

Table 5: Table of post-match standardized differences. Summary of data analyzed.

grouped by whether bail was set or not) to that in the matched setting (defendants grouped by whether they were arraigned by a strict or lenient judge). In doing so, we observe that the “near” aspect of near-far matching has improved covariate balance substantially, as standardized differences of the covariates decreased to nearly zero from the unmatched to matched settings. Similarly, in this plot, we see that the “far” aspect has also been achieved because the the standardized difference of the IV has *increased* from the unmatched to the matched analysis. By minimizing observed differences (”near” in covariates) and increasing differences in the IV (”far”) the designed study provides more evidence that variation in the outcome is due to the treatment.

Missingness

Next, we address the missingness in the outcome variable. Here, missingness arises because the case has yet to be resolved– sometimes because the defendant did not show up for their court date and sometimes because they have not yet plead guilty and are still awaiting trial. We find that 3.89% of the defendants in the encouraged group (i.e. those whose arraignments were overseen by stricter judges) had unresolved cases, whereas 4.09% of those in the unencouraged group did. These numbers are sufficiently similar that we conclude that there is no relevant difference between the encouraged and unencouraged groups in terms of their likelihood of having a missing value for their outcome. As such, for any pair in which at least one of the defendants in the pair had an unresolved case, we drop both records in the pair from our analysis. This leaves us with n = 56,734 individuals in the population of study. The process by which data is selected for inclusion in the analysis is summarized in Figure 3.

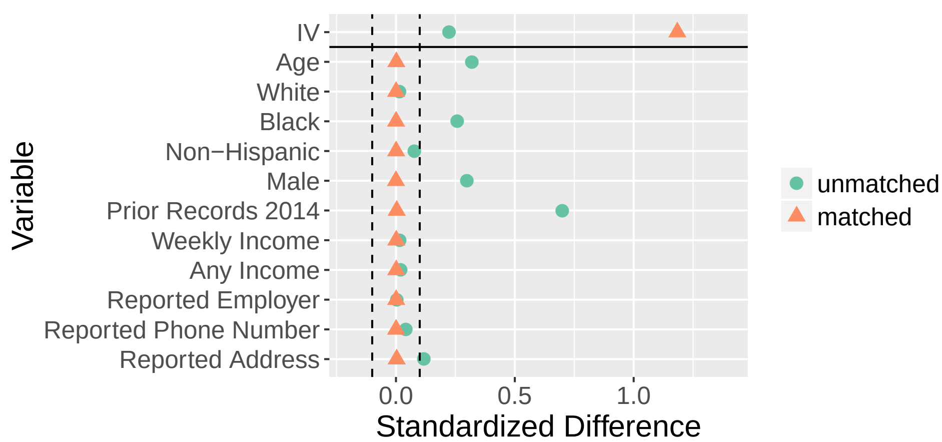


Figure 2: This graphic compares covariate balance between a naïve analysis applied to the full data (“unmatched”) and an analysis of the data using near-far matching (“matched”).

Generalizability

Because our methodology drops some participants from the study so that we can obtain optimal matching, the next question to address is whether our matched sample– the population from which we will make estimates– is informative about the full dataset, the clients of NYLAS in 2015. Figure 4 shows a side-by-side comparison of all binary variables for the full dataset and the matched dataset. Figure 5 shows comparisons for the remaining, non-binary variables. For almost all variables in this study, the distribution of the variable in the matched dataset is very similar to that of the full dataset. Acknowledging some minor differences, we believe that the magnitude of the differences between the matched population and the full dataset are generally small enough to not be substantively meaningful, thus we believe that the results from our matched group are generalizable to an analysis of the study population.

Estimates

Table 6 shows our estimates of λ, our measure of the causal impact of setting bail on the outcome of the case. The Estimate column displays a point estimate of λ. The Lower and Upper columns give the end points of a 95% confidence interval. The column labeled as n reports the number of observations in each stratum. The final column indicates whether the estimates are statistically significant at the $\alpha = 0.05$ level. We focus attention on the estimate at the top of Table 6 referred to as the total estimate in the aggregate stratum.

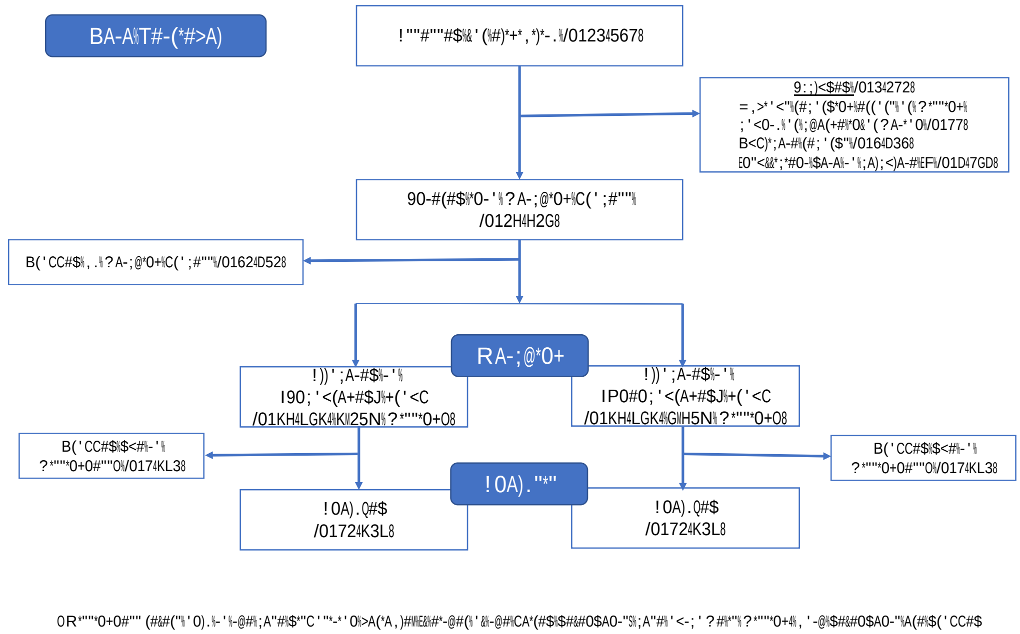


Figure 3: CONSORT flow diagram which shows the number of records dropped from the final analysis at each stage of the procedure.

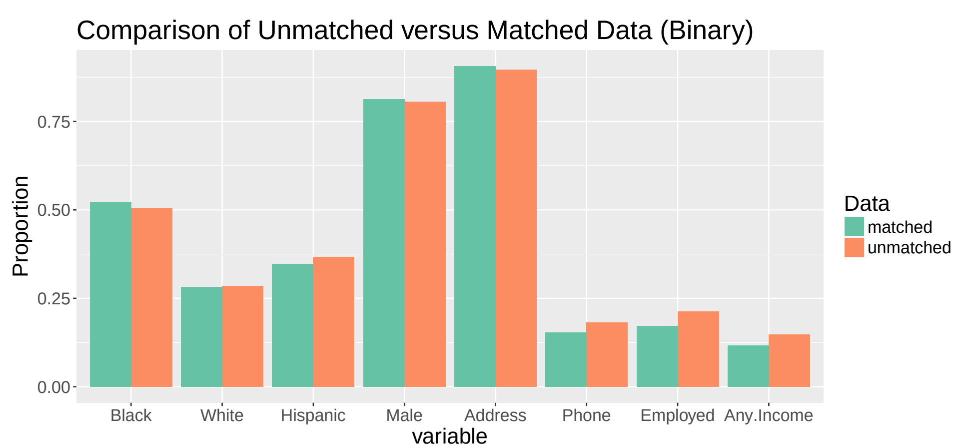


Figure 4: Comparisons of binary variables between full and matched datasets. Similar values in both datasets indicate generalizability of the analysis to the whole study population.

This is the global estimate across all case and defendant types and the focus of this study. This estimate should be interpreted as follows: for every additional 100 defendants that

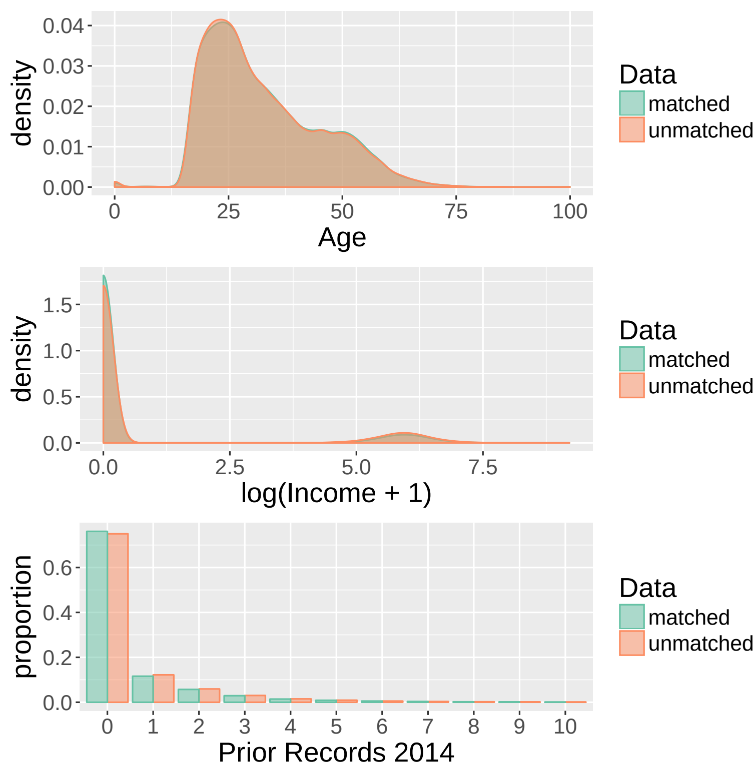


Figure 5: Comparisons of non-binary variables between full and matched datasets. Similar distributions in both datasets indicate generalizability of the analysis to the whole study population.

are assigned bail simply because they saw a stricter judge, an additional 34 guilty pleas or convictions will result that otherwise would not have. This represents a contextually meaningful increase in the probability of a guilty finding if bail is set.

We also present stratum-specific estimates for a variety of stratification schemes. The focus of our analysis is on the aggregate effect estimate, though we report the others for completeness. For many stratum-specific estimates, there is insufficient data to obtain estimates with small enough confidence intervals to definitively determine whether there was a positive or negative impact. In some cases, the absolute value of the end points of the confidence interval exceeds one, resulting in an estimated interval that extends beyond the possible range for an estimate that corresponds to an increase in probability. Although the interval contains out-of-bounds values, we report them as-is to emphasize the instability of those particular estimates. Due to the reduced sample sizes from stratifying and resulting reduction in statistical power, statistically significant differences between stratum-specific estimates are not possible. However, these stratum-specific estimates are suggestive sub-analyses that can be used to guide future research.

Stratification	Stratum	Estimate	Lower	Upper	n	Signif
Aggregate	total	0.34	0.2	0.49	56734	*
County	New York	0.43	0.23	0.63	17010	*
	Kings	0.34	0.14	0.54	17936	*
	Bronx	-0.07	-0.52	0.33	7290	
	Queens	0.66	0.13	1.35	12174	*
	Richmond	0.88	0.11	2.89	2324	*
Crime Type	Felony	0.22	-0.12	0.58	8448	
	Misdemeanor	0.37	0.22	0.53	48286	*
Gender	Male	0.31	0.16	0.45	46118	*
	Female	0.65	0.12	1.27	10532	*

Table 6: Estimated causal impact of setting bail on judicial outcome

Sensitivity Analysis

Inherent to any methodology that relies upon instrumental variables is an unverifiable assumption regarding the IV's relationship to the (unobserved) covariates, treatment, and outcome. If these assumptions are unmet, it is possible to estimate a causal relationship where none exists. We perform a sensitivity analysis to assess the robustness of our inference that there exists a positive causal relationship between setting bail and conviction in a case. The sensitivity analysis asks how unmeasured factors about the case might impact judge assignment and how this would alter our conclusions. For example, suppose there are two defendants with identical covariates, but one has a disheveled appearance and the other does not. Suppose that disheveled defendants are more likely to be assigned to stricter judges, and clean-cut defendants less likely. If disheveled defendants are also more likely to be convicted (simply because of their appearance), then because they will have bail set at a higher rate due to their higher likelihood of being assigned to stricter judges, we might erroneously infer that it was bail that caused the increased probability of conviction, not their appearance. Let Γ represent the difference in odds of assignment to a strict judge between a disheveled and clean-cut defendant. We can then test how large Γ can possibly be before we would find no statistically significant positive causal relationship between setting bail and conviction.

We use the method described in Baiocchi et al. (2010) to calculate the maximum such Γ and find that it is $\Gamma_{\max} = 1.064$. This method for calculating $\Gamma_{\max}$ is equivalent to that discussed in more depth in Small and Rosenbaum (2008). To put this in context using techniques described in Rosenbaum and Silber (2009), an unobserved covariate associated with a 50% increase in the odds of conviction and a 37% increase in the odds of being assigned to a strict judge corresponds to a $\Gamma_{\max}$ of 1.064.[4] Thus, in order for the inference that there exists a positive causal relationship to be false, it would have to be the case that there is some excluded variable that increases both one's odds of assignment to a strict judge and also one's odds of conviction. The magnitude of this increase would have to be

4. There are many pairs of values, {50%, 37%} that correspond to a $\Gamma_{\max} = 1.064$. We have chosen one such pair on which to focus the exposition because it seems reasonably plausible. Another example of a pair of values that corresponds to our $\Gamma_{\max}$ is {100%, 21%}

similar to increasing the odds of assignment to a strict judge by a third and increasing the odds of conviction by half.

In order to further contextualize whether $\Gamma_{\max}$ is convincingly large, we estimate the effect of several observed variables on the odds of conviction using a logistic regression with outcome variable G and covariates black, male, Hispanic, charge type, any income, reported employer, and reported phone number using the all of the matched data. Using the exponentiated coefficients as measures of the associated covariate's effect on the odds of conviction, we then calculate the necessary increase in odds of assignment to a strict judge to achieve $\Gamma_{\max} = 1.064$. For example, in the matched data, we find that the black covariate is associated with a 23% increase in the odds of conviction. An excluded variable with an effect on the odds of conviction as large as that of the black covariate would also have to increase the odds of assignment to a strict judge by 84% to achieve the estimated $\Gamma_{\max}$ in order to reverse the finding of a statistically significant positive causal relationship. Similarly, we find that reporting any income under our simple regression model is associated with a 34% increase in the odds of conviction. This may seem counter-intuitive as one may think that those who have income and are employed are less likely to commit a crime. However, this estimate is in line with the hypothesis that those who have jobs and are detained are more incentivized to accept a plea deal to return to their job with fewer days missed. Regardless, we find that in order to nullify our findings, we would need to be omitting a variable of similar importance to reported income that is also associated with a 55% increase in the odds of assignment to a strict judge.

The sensitivity analysis focuses on the potential for sorting to judges, a violation of one of the IV assumptions. What are some ways in which the sorting of cases with a higher likelihood of conviction to stricter arraignment judges might occur? One way would be if the arresting officer, or prosecutor, selected which judge oversaw the arraignment. These officials might then assign individuals whom they believed had a higher likelihood of conviction to the strictest judge on duty. As described in section 2, defendants are arraigned by whichever judge is on duty when their turn comes, and there is no human discretion involved in routing defendants to judges. Thus this is unlikely the case. Another way this sorting might occur is if there were systematic correlations between police activity and judge scheduling. For example, suppose the strictest judges always work on Tuesdays. If the police also rounded up particularly egregious cases beginning Monday night so that they would be arraigned by a judge on Tuesday, then this could cause systematic sorting of more serious cases to strict judges. Yet even if this occurred, it would have to be the case that they rounded up especially heinous cases with a particular top charge on Monday nights relative to all other cases with that top charge. For example, it would not be enough that they round up murder suspects on Monday nights; they would have to round up suspects of especially egregious murder cases on Monday nights. To our knowledge, it is not the case that suspects of especially serious variants of a top charge would be arrested in such a way that those cases would be synchronized to the scheduling of stricter judges. Regardless, if some such mechanism does exist, in order to nullify our findings, it would have to result in charges on the order of a 37% increase in the likelihood of being arraigned by a strict judge and a 50% increase in the odds of conviction. We find this unlikely.

There is a second assumption inherent to our IV that is not addressed in this sensitivity analysis: that the judge only affects the final disposition of the case through her decision to set bail. As discussed in the introduction, the arraignment judge has no role in the handling of the case post-arraignment. However, it is possible that in addition to issuing a bail decision, the judge also makes comments that influence the prosecutor in how they handle the case going forward or to the defendant that alter their perception of the fairness of the system or their chances of being acquitted should the case go to trial. If this is true, the judge assignment would affect the case outcome through some mechanism other than bail setting and the likely resultant pre-trial detention. We expect that if such an effect does exist, it would be very small relative to the effect of bail. While we cannot prove that this effect is negligible or non-existent, it is worth noting that the assumption that judges only affect case disposition through the setting of bail is consistent with other studies in this area.

6. Discussion

The work we have presented adds to the mounting empirical evidence that money bail and the pre-trial detention it precedes causes a higher likelihood of conviction. Though we approach the problem using a different tradition for analyzing observational data than other similar studies, our substantive findings support the conclusions of the recent literature in this area. We find a strong causal relationship between setting bail and the outcome of a case for the clients of NYLAS– specifically, we find that for cases for which different judges could come to different decisions regarding whether bail should be set, setting bail results in a 34% increase in the chances that they will be found guilty.

A substantively meaningfully causal impact of bail on case outcome calls into question the fairness of the practice. These results suggest that the money bail system creates a situation where the odds of conviction can be materially increased due to a hasty, perhaps seemingly inconsequential decision that may have been be made before all relevant evidence could be compiled. Contextual information about the causal mechanism at work implies that setting bail causes a higher likelihood of conviction specifically by acting as a barrier to pre-trial release. That our estimate is higher than reported in other recent work is consistent with this causal mechanism, as this would imply that the barrier to release would be higher for those who are less able to afford a given amount of bail (such as those who rely on public defenders) causing a larger effect. It is also likely that our estimate deviates from other recently reported estimates because of how we define the pertinent population. For example, one of the recent cited studies considered only felonies and most consider different time periods and locations. Though it is not explicitly mentioned, it seems that several studies include cases that were disposed at arraignment, whereas we define our population to be cases that have made it past that stage. Without more detailed information about the data and the pre-processing steps taken prior to the main analyses presented elsewhere, it is difficult to assess how our study population differs from others and whether the discrepancy is due to differences in methodology or population definitions. We believe that our study contributes to this field by clearly laying out all of these steps for future comparison. Regardless of the precise magnitude of the effect, all comparable studies including the work we have presented find a strong positive causal relationship. The real world implications of this are that there are likely many people– disproportionately, poor people– who have been convicted of crimes simply because bail was set.

Appendix A. Results from excluding ACD cases

Here, we present analogous results to those provided above, having excluded cases that received an ACD at arraignment. The initial sample size without ACDs is 59,571. After matching, we are left with 46,840 cases.

In general, we have substantively similar findings, though the reduced power that comes with a smaller sample size means that in some cases, previously significant findings are no longer statistically significant. Table 7 shows the standardized differences after matching the non-ACD cases. We again find excellent balance among the encouraged and unencouraged groups.

	Strict n=21525	Lenient n=21525	Abs St Dif
Guilty	0.55	0.54	0.03
Bail Set	0.28	0.22	0.13
IV	-0.09	0.08	1.28
Age	34.22	34.27	0.00
White	0.28	0.28	0.00
Black	0.55	0.55	0.00
Non-Hispanic	0.66	0.66	0.00
Male	0.83	0.83	0.00
Prior Records 2014	0.65	0.66	0.00
Weekly Income	63.60	62.53	0.00
Any Income	0.14	0.14	0.00
Reported Employer	0.20	0.20	0.00
Reported Phone Number	0.17	0.17	0.00
Reported Address	0.92	0.92	0.00

Table 7: Table of post-match standardized differences. Summary of data analyzed.

Figure 6 shows a comparison of covariate balance before and after matching.

Table 8 shows our estimated effects, this time without including ACDs. This is the analogous table to table 6 in the main results.

Stratification	Stratum	Estimate	Lower	Upper	n	Signif
Aggregate	total	0.24	0.1	0.38	43050	*
County	New York	0.37	0.18	0.56	12640	*
	Kings	0.21	0.01	0.41	13620	*
	Bronx	-0.03	-0.47	0.36	5496	
	Queens	0.55	-0.09	1.43	9132	
	Richmond	0.97	0.37	2.3	2162	*
Crime Type	Felony	0.29	-0.02	0.64	8634	
	Misdemeanor	0.23	0.07	0.39	34416	*
Gender	Male	0.24	0.1	0.39	35846	*
	Female	0.19	-0.42	0.78	7174	

Table 8: Estimated causal impact of setting bail on judicial outcome

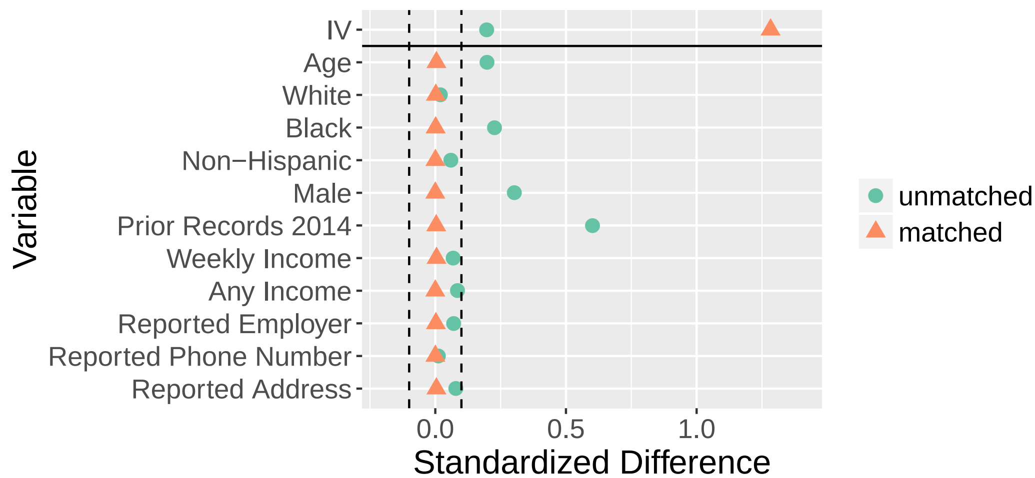


Figure 6: This graphic compares covariate balance between a naive analysis applied to the full data ("unmatched") and an analysis of the data using near-far matching ("matched"). For details on how this figure diverges in interpretation from the standard Love plot, see section 5 of the main text.

Graphics to assess the generalizability of these results are shown in Figures 7 and 8.

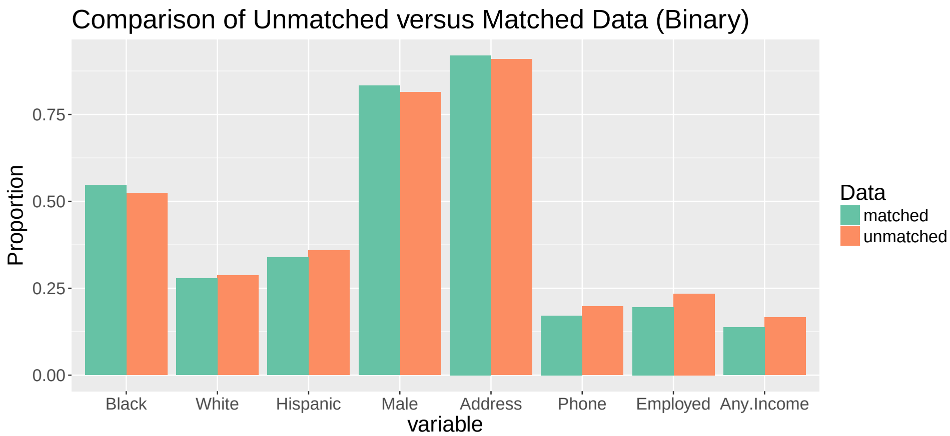


Figure 7: Comparison of binary covariates between matched dataset used in analysis and full, unmatched dataset.

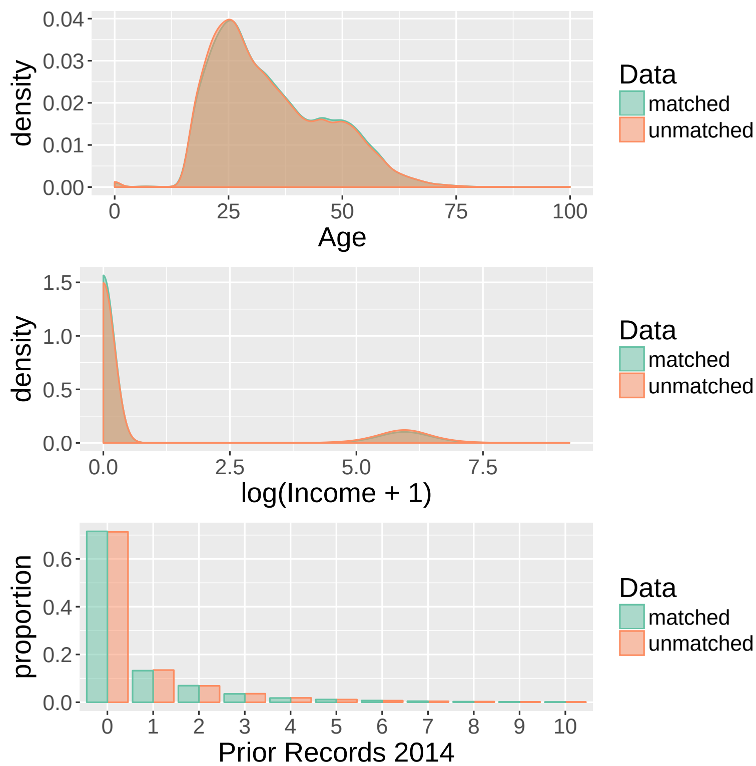


Figure 8: Comparison of non-binary covariates between matched dataset used in analysis and full, unmatched dataset.

References

Aizer, A. and Doyle, J. J. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, page qjv003.

Ares, C. E., Rankin, A., and Sturz, H. (1963). The Manhattan Bail Project: An interim report on the use of pre-trial parole. *NYUL rev.*, 38:67.

Baiocchi, M., Small, D. S., Lorch, S., and Rosenbaum, P. R. (2010). Building a stronger instrument in an observational study of perinatal care for premature infants. *Journal of the American Statistical Association*, 105(492):1285–1296.

Baiocchi, M., Small, D. S., Yang, L., Polsky, D., and Groeneveld, P. W. (2012). Near/far matching: a study design approach to instrumental variables. *Health Services and Outcomes Research Methodology*, 12(4):237–253.

Berube, D. A. and Green, D. P. (2007). The effects of sentencing on recidivism: Results from a natural experiment.

Bharara, P. (2014). CRIPA investigation of the New York City Department of Correction jails on Rikers Island.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, 90(430):443–450.

Danziger, S., Levav, J., and Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17):6889–6892.

Dexter, P. R., Wolinsky, F. D., Gramelspacher, G. P., Zhou, X.-H., Eckert, G. J., Waisburd, M., and Tierney, W. M. (1998). Effectiveness of computer-generated reminders for increasing discussions about advance directives and completion of advance directive formsa randomized, controlled trial. *Annals of internal medicine*, 128(2):102–110.

Dobbie, W., Goldin, J., and Yang, C. (2016). The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. Technical report, National Bureau of Economic Research.

Eren, O. and Mocan, N. (2016). Emotional judges and unlucky juveniles. Technical report, National Bureau of Economic Research.

Green, D. P. and Winik, D. (2010). Using random judge assignments to estimate the effects of incarceration and probation on recidivism among drug offenders. *Criminology*, 48(2):357–387.

Gupta, A., Hansman, C., and Frenchman, E. (2016). The heavy costs of high bail: Evidence from judge randomization. *The Journal of Legal Studies*, 45(2):471–505.

Hickam, D., Totten, A., Berg, A., Rader, K., Goodman, S., and Newhouse, R. (2013). The PCORI methodology report. *Washington: PCORI*.

Holland, P. W. (1988). Causal inference, path analysis and recursive structural equations models. *ETS Research Report Series*, 1988(1).

Human Rights Watch (2017). Not in it for justice: How california's pretrial detention and bail system unfairly punishes poor people. Technical report, Human Rights Watch.

Imbens, G. W. and Rosenbaum, P. R. (2005). Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1):109–126.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2017). Human decisions and machine predictions. *NBER Working Paper Weries*, (Working Paper 23180).

Kling, J. R. (2006). Incarceration length, employment, and earnings. *The American economic review*, 96(3):863–876.

Leslie, E. and Pope, N. G. (2016). The unintended impact of pretrial detention on case outcomes: Evidence from nyc arraignments. *Unpublished Working Paper*.

Lorch, S. A., Baiocchi, M., Ahlberg, C. E., and Small, D. S. (2012). The differential impact of delivery hospital on the outcomes of premature infants. *Pediatrics*, pages peds–2011.

Lowenkamp, C. T., VanNostrand, M., and Holsinger, A. (2013). The hidden costs of pretrial detention. *Laura and John Arnold Foundation.*

Lu, B., Zanutto, E., Hornik, R., and Rosenbaum, P. R. (2001). Matching with doses in an observational study of a media campaign against drug abuse. *Journal of the American Statistical Association*, 96(456):1245–1253.

Martin, S. E., Annan, S., and Forst, B. (1993). The special deterrent effects of a jail sanction on first-time drunk drivers: A quasi-experimental study. *Accident Analysis & Prevention*, 25(5):561–568.

Phillips, M. T. (2007a). Bail, detention, and nonfelony case outcomes. *Research Brief.*

Phillips, M. T. (2007b). *Pretrial Detention and Case Outcomes, Part 1: Nonfelony Cases.* CJA, New York City Criminal Justice Agency, Incorporated.

Phillips, M. T. (2008). *Pretrial Detention and Case Outcomes, Part 2: Felony cases.* CJA, New York City Criminal Justice Agency, Incorporated.

Rankin, A. (1964). The effect of pretrial detention. *NYUL Rev.*, 39:641.

Rigdon, J., Baiocchi, M., and Basu, S. (2016). nearfar: Near-Far Matching. *R package version*, 1.0.

Rosenbaum, P. R. and Silber, J. H. (2009). Amplification of sensitivity analysis in matched observational studies. *Journal of the American Statistical Association*, 104(488):1398–1405.

Silber, J. H., Rosenbaum, P. R., Trudeau, M. E., Even-Shoshan, O., Chen, W., Zhang, X., and Mosher, R. E. (2001). Multivariate matching and bias reduction in the surgical outcomes study. *Medical care*, pages 1048–1064.

Small, D. S. and Rosenbaum, P. R. (2008). War and wages: the strength of instrumental variables and their sensitivity to unobserved biases. *Journal of the American Statistical Association*, 103(483):924–933.

Stevenson, M. (2016). Distortion of justice: How the inability to pay bail affects case outcomes.

Warren, G., Lamont, I., and of America, U. S. (1972). The unconstitutional administration of bail: Bellamy v. the judges of New York City. *Criminal Law Bulletin*, 8(6):459–506.

Weinshall-Margel, K. and Shapard, J. (2011). Overlooked factors in the analysis of parole decisions. *Proceedings of the National Academy of Sciences*, 108(42):E833–E833.

Zubizarreta, J. R., Small, D. S., Rosenbaum, P. R., et al. (2014). Isolation in the construction of natural experiments. *The Annals of Applied Statistics*, 8(4):2096–2121.