Make-or-Buy? The Provision of Indigent Defense Services in the U.S.*

Yotam Shem-Tov

(UC Berkeley)

Abstract

U.S. courts provide constitutionally mandated legal services to low-income defendants via private court-appointed attorneys and public defenders' organizations. This paper investigates the relative efficacy of these two modes of indigent defense by comparing outcomes of co-defendants assigned either a public defender or a private court-appointed attorney within the same case. Using data from San Francisco and federal district courts, I find that in multiple defendant cases public defender assignment is plausibly as good as random. Public defenders reduce the probability of any prison sentence by 22% and the length of prison by 10%.

Keywords: indigent defense, criminal justice, crime, provision of public services

JEL: H44, K14, K42, J15

*Yotam Shem-Tov: Department of Economics, University of California Berkeley, 530 Evans Hall, Berkeley, CA 94720, USA; shemtov@berkeley.edu. Previous versions of this paper were entitled "An Investigation of Indigent Defense Systems: Public Defenders vs. Court Appointed Private Attorneys." and "Public Defenders vs. Court Appointed Private Attorneys: An Investigation of Indigent Defense Systems." I am grateful to David Card for all his support, guidance, and comments. I thank Jeff Adachi, Tyler Vu, and Simin Shamji from the San Francisco Public Defender Office for numerous talks and discussions. I thank Alan Auerbach, Alma Cohen, Paul Hofer, Patrick Kline, Justin McCrary, Conrad Miller, Kevin Quinn, Ilyana Kuziemko, Steven Raphael, Evan Rose, Emmanuel Saez, James Sallee, Jasjeet Sekhon, Christopher Walters, Nicholas Li, Juliana Londoño-Vélez, Maxim Massenkoff, and participants in the 11th Annual Conference on Empirical Legal Studies, the UC Berkeley Public Finance Lunch Seminar, the UC Berkeley Labor Lunch Seminar, and the 27th Annual Meeting of the American Law and Economics Association for helpful comments and suggestions. I gratefully acknowledge financial support from the U.S. Bureau of Justice Statistics and the Center for Equitable Growth.

"Even the intelligent and educated layman... requires the guiding hand of counsel at every step in the proceedings against him. Without it, though he be not guilty, he faces the danger of conviction because he does not know how to establish his innocence."
Justice George Sutherland, 1932.

I Introduction

Indigent defendants facing criminal charges in both state and federal courts in the U.S. have a constitutionally protected right to legal counsel from an attorney who is appointed and compensated by the state. In the U.S., 80% of criminal defendants require the assistance of indigent defense services. Legal counsel is essential for a defendant to receive a fair trial because "the average defendant does not have the professional legal skill to protect himself when brought before a tribunal with power to take his life or liberty" (Johnson v. Zerbst, 1938). There are two common alternatives for providing this public service: public defender organization (henceforth PD) and court-appointed private attorneys (henceforth CA).[1] Any systematic gap in the quality of legal representation that defendants receive, which depends on the type of legal counsel to whom they are assigned—PD or CA—raises concerns regarding violations of the defendants' constitutional Sixth Amendment rights (Joy and Mcmunigal, 2012).[2]

Differences across attorney types in case outcomes can have long-lasting impacts on defendants and their families. In addition to the risk of being sent to prison, involvement in the criminal justice system has a negative impact on earnings, employment, and educational attainment (Dobbie et al., 2016; Mueller-Smith, 2015; Aizer and Doyle, 2015; Raphael, 2011). While the impact of judges (Yang, 2015; Cohen and Yang, 2018; Abrams et al., 2012), prosecutors (Rehavi and Starr, 2014) and parole boards (Kuziemko, 2013) has been documented in the literature, the importance of the defense attorney—especially among indigent defendants—has received less attention.

This paper investigates the relative efficacy of these two modes of providing legal counsel to low-income individuals. The main challenge in evaluating the performance of PD relative to CA is that the usual mechanism of assigning an indigent (i.e., low-income) defendant to a PD or a CA is *not* random and can vary across jurisdictions. While defendants cannot manipulate the process, the judge, court, and public defender's office can potentially influence the assignment procedure. I

[1] In state courts, CAs are also referred to as conflict attorneys, assigned counsel, or panel attorneys. In federal courts, CA are commonly referred to as CJA attorneys in reference to the Criminal Justice Act of 1964, which established the federal indigent defense system. In this paper, I refer to court-appointed private attorneys as CA and use the initials CJA as a reference to the Criminal Justice Act of 1964.

[2] Today, CAs are common in indigent defense systems as either a substitute to a PD office or as a complement, handling cases the PD cannot represent due to conflicts of interest. In the last fifty years, there has been an increase in the number of counties and states that operate a PD organization as part of their indigent defense system. Still, across the U.S., there is heterogeneity in the types and structures of indigent defense systems (DeFrances and Litras, 2000; Farole and Langton, 2010a,b). Appendix (C) provides a review of the history of the right to appointed-counsel in the U.S. at both the state and the federal court systems.

use administrative court records from two sources. All cases terminated in San Francisco between 2006 to 2016, and all federal district court cases terminated between 1996 and 2014. I find extensive evidence of non-random sorting of defendants across PD and CA in both San Francisco and federal district courts. This selection confounds a causal interpretation of comparisons of case outcomes between defendants represented by a PD and a CA.

This paper employs a new identification strategy of comparing co-defendants within the same case. In multiple defendant cases the PD office does not represent co-defendants to avoid inherent conflicts of interest (Moore, 1984; Allison, 1976; Lowenthal, 1978). In general, the within-case assignment of defendants to a PD does not have to be random; however, I show that in San Francisco, the decision of who will be assigned a PD in multiple defendant cases can plausibly be treated as good as random. The within-case assignment to a PD is not correlated with defendant characteristics such as race, age, criminal history, and charge severity. Selection on unobserved factors is unlikely, since these omitted variables need to be correlated with both case outcomes and PD assignment, but uncorrelated with criminal history, charge severity, age, and race. I exploit this natural experiment to quantify the causal effect of being assigned a PD relative to a CA on case outcomes. In contrast, in federal courts, I document both across and within-case selection. Within a multiple defendant case, the order in which defendants are listed on the indictment is correlated with both PD assignment and the defendant's culpability. To overcome this issue, I condition on the defendant's order of appearance on the indictment using an order-specific fixed effect. Once the defendant's position on the indictment (e.g., first, third) is taken into account, the assignment to a PD or a CA can be treated as if it were done independently of the defendant's culpability.

I find that defendants assigned to a PD generally obtain more favorable sentencing outcomes both in San Francisco and in federal courts; however the magnitude of the effects is smaller in federal courts. In San Francisco, defendants assigned to a PD have a lower probability of both conviction (6.4%), and prison sentence (22%), as well as a shorter expected imprisonment term (10.8%). There is no evidence of heterogeneity with respect to the defendant's demographic characteristics (e.g., gender, race); however, there is substantial heterogeneity with respect to criminal history and charge severity: defendants who face more severe criminal charges (e.g., felony vs. misdemeanor) and have a longer criminal history are the ones driving the results. This implies that as the likelihood of incarceration is higher, due to having a longer criminal history or facing more severe charges, the effect of having a better legal counsel will be larger. The estimated attorney type effects are not correlated with the number of multiple defendants in the case.

In federal district courts, the findings are qualitatively similar, although the magnitude of the effects is much smaller. Defendants assigned a PD face a shorter prison sentence by 4.64% and a lower probability of incarceration by 1%, yet the probability of conviction is not influenced by the attorney type. The vast majority of crimes are prosecuted at state courts (e.g., San Francisco) and differences between the state and federal systems are important for understanding the external validity of the results in this study and others (Iyengar, 2007; Anderson and Heaton, 2012; Roach, 2014). The overall rates of conviction and incarceration differ dramatically between San Francisco and federal courts. For example, in federal courts 95% of defendants are convicted and 86% are sentenced to prison relative to 59.8% and 6.3% in San Francisco. These differences in rates can explain why being assigned a PD has a lower impact in federal courts since the overall chances not to be incarcerated or convicted are small.[3]

Having established that PDs provide better legal representation, I turn to investigate how much of these results can be explained by differences in attorney observable characteristics using data from San Francisco. Difference in these characteristics provide descriptive evidence about the adverse selection into the pool of attorneys who choose to accept indigent defense appointments relative to the attorneys who choose to work for a PD office. I find that PDs are younger, demographically more diverse (higher share of females and non-whites), graduate from B.A. and J.D. programs in higher-ranked institutions, and have more court experience.[4] I find that attorney characteristics account for slightly less than half of the estimated attorney type effects on the probability of conviction or prison sentence; and the sentenced prison length. Furthermore, the estimated differences in sentencing outcomes have the same sign as before, but are not statistically significant. These findings suggest that adverse selection of attorneys to a PD office relative to CA is the key driver of the differences in case outcomes.

This paper contributes to a large literature on whether the state should "make-or-buy" public services. Other examples of such decisions range from schools (Abdulkadiroglu et al., 2015) to police (Cheng and Long, 2017) and prison (Mukherjee, 2017).[5] Weak populations (e.g., prisoners, criminal defendants) are especially vulnerable to the privatization of public services (Hart et al., 1997). More closely related to this study, is recent work that utilized various empirical methods to evaluate the efficacy of PD relative to CA (Iyengar, 2007; Anderson and Heaton, 2012; Roach, 2014).

Anderson and Heaton (2012) exploit the initial random assignment of defendants in murder cases in Philadelphia to CA and PD to compare between the two. They find that being assigned a PD reduces the defendant's sentenced imprisonment time by 31% but has *no* effect on the probability of conviction. The present paper extends Anderson and Heaton (2012) in several

[3] In San Francisco, the share of defendants sentenced to prison is 6.3% and the share who are sentenced to either jail or prison is 16%.

[4] Differences in the financial incentives facing the two attorney types may also play a role as PDs are salaried workers and CAs are compensated on a per-case basis. Two related papers discuss the impact of moral hazard on CAs. Agan et al. (2017) compare CAs to privately retained attorneys and find that wage incentives influence the defendants' case outcomes. They employ an order-invariant decomposition method that was proposed by Gelbach (2016). Schwall (2015) also investigates the effect of wage incentives on the behavior of CAs by leveraging administrative records from South Carolina and a reform that changed their compensation schedule from an hourly wage rate to a flat fee. He finds that under the flat fee regime attorneys report significantly less hours worked; however, he finds no impact on the defendants' case outcomes (e.g., conviction, prison sentence). Both of this papers study CAs and do not make a comparison between PDs and CAs.

[5] The term "make-or-buy" was coined by Coase (1937) and is often applied to how firms should carry out their manufacturing processes.

directions. The first is external validity. I evaluate PDs and CAs in a range of different offenses and not only in capital murder trials, which are a rare and not representative procedure, constituting 0.1% of arrests in the U.S. in 2016.[6] In San Francisco, I find that assignment to a PD causes a significant reduction in the probability of conviction, unlike the Anderson and Heaton's results. I also find that the majority of the differences in prison sentences between attorney types is driven by felony cases and defendants with a prior criminal history who face a higher probability of incarceration. Since capital murder is the most severe offense, it is expected that the attorney type effects will be the largest in these cases, which also explains why my estimates are lower (a 10.8% relative to a 31% reduction in incarceration length). These results are important for understanding at which situations there is (or is not) a gap in case outcomes between PD and CA and show that for less severe offenses than murder the attorney type can impact the probability of conviction and not only the length of incarceration. Second, I conduct a comparison between the relative efficacy of PDs in the state and federal court levels and find consistent estimates, although there are differences in magnitudes. This comparison is important for understanding how informative are estimates from one system about the other. Third, I quantify how much of the estimated gaps can be accounted for by attorney-observable characteristics (approximately 50%); and show that selection of attorneys into PD relative to CA can account for most of the differences in efficacy. Fourth, I evaluate whether assignment to a PD leads individuals to re-offend more or less in the future, i.e., whether individuals who received more lenient case outcomes because of the attorney type are more or less likely to re-offend. I find that defendants who have been assigned a PD are more likely to recidivate; however, the differences are not statistically significant.

Selection in the assignment of defendants across attorney types, coupled with the fact that most jurisdictions do not explicitly randomly assign defendants across attorney types, suggests studies that naïvely compare these two attorney types will tend to overestimate the efficacy of PDs relative to CA. For example, in San Francisco, the effect of being assigned a PD on prison length changes from -35.7% to -18% when comparing across cases. This is evidence that individuals who are likely—based on observable—to receive a shorter prison sentenced are the ones who are allocated to a PD. Unlike the *across* cases comparison, in the within-case comparison the estimates with and without controls are similar, -10.8% and 10.9%; and are *lower* in magnitude.

Iyengar (2007) and Roach (2014) both argue that by using a data-driven procedure it is possible to detect location-year pairs in which the attorney type assignment was done at random. The methodology is based on examining the relationship between observed defendant characteristics and attorney type assignment within a specific location-year pair. Assuming locations can randomize in *select* time periods, they examine whether a location in a given period randomized defendants to attorney types by conducting a joint F-test to determine whether case and defendant

[6] See https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/topic-pages/tables/table-21 for the share of arrests that are for murder offenses across the U.S. according to the FBI's Uniform Crime Report.

characteristics predict PD assignment. I examine the performance of this data-driven procedure in federal courts and find two indications that it causes false classifications that can bias the estimated effects. First, districts do not pass the randomization test for several *consecutive* years. To illustrate, consider a hypothetical district A that passes the test in 2001 but not in 2002 nor in subsequent years, but does pass the test again in 2009. This is puzzling, as districts are unlikely to change their method of allocation so often. Second, once the assessment of covariate balance is done by pooling all district-year pairs that passed the F-test, including district and year FEs, I find significant covariate *imbalances* in charge severity measures across defendants who are assigned a PD relative to a CA.[7] My estimates of assignment to a PD in this paper are substantially different in magnitude than the ones reported by both Iyengar (2007) and Roach (2014). One explanation for the differences is that the above data-driven procedure does not correctly assess whether an allocation is truly random.

My findings regarding differences in attorney quality inform the policy discussion about the high rates of incarceration in the U.S. For instance, can we reduce the share of the population that is behind bars by providing better legal assistance? My estimates indicate that assignment to a PD decreases the likelihood of being sentenced to prison by 22% in state courts, but has a negligible impact in federal courts 1%. This suggests that, at least in state courts, changing the method of provision can have a lasting impact on the share of defendants who are incarcerated as well as on re-offending patterns and thus long-term relationships with the criminal justice system. Better legal representation can also cause an increase in crime if more lenient case outcomes enable defendants to re-offend easier or sooner (Ater et al., 2017).

The remainder of the paper is organized as follows. Section (II) describes the data. Section (III) presents the empirical framework. Section (IV) describes the identification strategy and verifies the validity of the multiple defendant design. Section (V) presents the empirical findings. Section (VI) concludes and briefly suggests avenues for future research.

II Data

II.A Data sources and sample construction

San Francisco

I use administrative records from the court system in San Francisco for all cases terminated between February 2006 and March 2016. The data contains sentencing outcomes such as conviction, length of prison sentence, and length of probation, as well as a detailed description of the filed charges ranging from broad characteristics such as felony classification to more granular information on

[7]Note that, Iyengar (2007) and Roach (2014) both combine together all the location-year pairs that pass the $F - stat$ test and include location and year FEs in their main outcome analysis. Appendix B describes the implementation of the data-driven procedure in federal district courts and the covariate balance analysis.

the specific statute and title of the offense. I also calculated the SC and BCS codes for each charge, which are classification of offenses to broader categories. Basic demographic information on the defendant such as race, sex, and age is available, and I use names to infer Hispanic origin using data from the 2000 Census.[8]

Defendants often face multiple charges at the time of disposition for offenses that took place at different times. For example, an individual can be charged with offense A and then be released on bail; while awaiting trial he can then be additionally charged with offense B. The disposition of both charges can take place at the same time. In the above scenario, if the defendant is indigent, the attorney who represented him for charge A will also be his counsel for charge B. Therefore, attorney assignments are based on the initial charging of each case. I group charges together into cases based on whether the conviction, offense, or charging date fall within a certain time window (e.g., 20 days). I then define the initial attorney type assignment as the first attorney that represented the defendant within a case.[9] For the main analysis, I restrict attention to criminal cases in which the defendant was initially represented by an appointed counsel: either a PD or a CA.

II.B Federal courts

The "Federal Court Cases: Integrated Database" is constructed by the Bureau of Justice Statistics and made available by the National Archive of Criminal Justice Data and the Inter-University Consortium for Political and Social Research at the University of Michigan. The data series covers every criminal case, in federal district courts, that was terminated from 1970 to 2014. It contains rich information on filed charges, case disposition, and sentencing outcomes. From 1996 onward, both the initial and final attorney types are available. Prior to 1996, only the attorney type at the time of disposition was recorded. For this reason, in the main analysis only cases that have been terminated/disposed from 1996 onward are used.

II.C Data descriptives

II.C.1 Defendants in San Francisco

The public defender's office in San Francisco was established in 1921 and represents the majority of indigent defendants. The CA attorneys in San Francisco (known as "conflict attorneys"), are considered to be professionals who provide competent legal counsel to their clients. They are not obliged to represent clients and the court compensates them for their work. CA attorneys must satisfy strict requirements to be eligible to receive indigent defense appointments from the court.

[8] SC and BCS codes are classification of offenses to broader categories. The classification is done by the California Department of Justice, https://oag.ca.gov/law/code-tables.

[9] The choice of using no time window or a 5, 10 or 20 days time window has no impact on the results. The effects look similar regardless of the choice.

Indigent defendants are generally assigned to the PD office in San Francisco unless there is a conflict of interest; only then are cases assigned to CAs.[10] Figure (1) shows the distribution of defendants across attorney types in San Francisco from 2006 to 2015. In single defendant cases (panel a), the vast majority of indigent defense representation is done by the PD office; however, within multiple defendant cases the division of is almost equal (panel b). This prevalence of CAs in multiple defendant cases results from the fact that the PD office in San Francisco avoids representing more than one defendant within a case. Interestingly, the share of defendants represented by a private attorney is the same in single and multiple defendant cases.

Table (1) presents descriptive information on criminal defendants in San Francisco. Column (2) includes all cases with more than one defendant and column (3) all cases with at lease one defendant that is represented by a PD and another by a CA such that both a PD and a CA are present in the case. Approximately 50% of the defendants are Caucasian and African-Americans are overly represented.[11] The share of African-Americans and females is higher in multiple defendant cases (columns 2 and 3) relative to single defendant cases. The average age in multiple defendant cases is lower than in single defendant cases 32 relative to 35. Multiple and single defendant cases vary also in the severity of the charges: 82.9% include a felony charge relative to 51.8% respectively; and the probability to be incarcerated in prison (jail) is higher (lower) in multiple relative to single defendant cases. In almost a quarter of the cases the charges are eventually dropped. Multiple defendant cases that include both a PD and a CA (column 3) are the majority of multiple defendant cases and are similar to the overall sample of multiple defendant cases (column 2) in defendant demographics, charge severity measures and case outcomes.

Defendants in federal district courts

Like defendants in San Francisco, the vast majority of federal defendants are represented by indigent defense services. The provision of these services is laid out by the CJA and its guidelines are set by the Judicial Conference of the United States.[12] Tables (2) and (A.1) show descriptive information of defendants in federal district courts for cases terminated between 1996 and 2014 and between 1970 and 1995. Cases in federal district courts usually end in a conviction and almost always through a plea bargain. Defendants in a multiple defendant case, on average, face more severe charges and their cases terminate with longer prison sentences, which puts them in crucial need of legal counsel.

[10] Conflicts of interest can occur under various circumstances. For example, in a multiple defendant case, the interests of the different individuals can contradict each other and they will each require a separate legal counsel. Another example is the attack that took place in Charlottesville, Virginia (link to article below). The PD office could not represent the accused in the attack since certain members of the office had family members who were wounded in the assault, and the defendant was assigned a CA. https://www.theguardian.com/us-news/2017/aug/14/james-fields-charlottesville-driver-murder-charge?CMP=Share_AndroidApp_Gmail.

[11] The share of African-Americans in the population of San Francisco was approximately 6% in 2010.

[12] The guidelines for the administration of the Criminal Justice Act are described online at: http://www.uscourts.gov/rules-policies/judiciary-policies/criminal-justice-act-cja-guidelines

The number of federal defendants has increased consistently between 1996 and 2014, and the proportion of indigent defendants has increased as well. Figure (A.1) reports the share of defendants represented by a PD, a CA, and/or a privately retained counsel. The share of defendants that are represented by PDs has also increased continuously over time and the share of defendants that retained a private counsel shows a downward trend. In 1970 approximately 40% of all federal defendants retained a private counsel, and in 2014 less than 20% did. The increasing share of defendants who cannot hire a private attorney highlights the importance of studying the provision of indigent defense services.

The number of federal PD organizations has steadily increased since the establishment of the federal PD program in 1970. In 1993, 48 PD organizations served 54 of the 94 federal districts (Prado et al., 1993), and as of 2016, 81 PD organizations have provided indigent defense services to 91 of the 94 federal districts. Although the Judicial Conference called "for the appointment of public defenders in those districts in which the amount of criminal litigation justified the presence of such an office" (Prado et al., 1993), the provision of indigent defense services using PD organizations started only in 1970 and before that time only CAs represented indigent defendants.

III Econometric framework

The main objective of this paper is to measure differences in quality of legal representation by estimating the causal effect of assignment to a PD relative to a CA on the case outcomes of the defendant. I argue that by conditioning on a sample of multiple defendant cases with both a PD and a CA within a case, the assignment to a PD can plausibly be considered as good as random. As I outline in further detail in Section (IV), a PD office is constrained to representing only one client in a multiple defendant case due to potentially conflicting interests between co-defendants. In Section (IV), the balance tests show that within a case, the defendants with a PD and those with a CA are not observably different in San Francisco. Comparing outcomes using *within-case* variation limits selection biases that may arise from comparing outcomes *between cases*.

Let $\text{PD}_i \in \{0, 1\}$ be an indicator of whether defendant i was first assigned a PD, and let Y_i denote some sentencing outcome of interest (e.g., length of imprisonment). A naïve causal model that relates the defendant's attorney type to his case outcomes is:

$$Y_i = \beta \cdot \text{PD}_i + X_i'\Gamma + \alpha_{j(i)} + \epsilon_{1i}, \tag{1}$$

where $j(i)$ is a mapping from defendant i to court case number j, X_i is a vector of observable pre-trial variables that include measures of the severity of the filed charges (e.g., offense codes) and the type of charges (e.g., felony, misdemeanor), the demographic characteristics of the defendants and their criminal history; and β is the effect of assignment to a PD on case outcome Y_i.

Anecdotes from federal public defenders suggest that in some federal districts, defendants are sorted on the indictment in the order that law enforcement officials think reflects culpability. Perhaps intentionally, in federal district courts the order of the defendants on the indictment is correlated with assignments to a PD. To account for this, I consider a second causal model that flexibly controls for the order of appearance of the defendants on the indictment:

$$Y_i = \beta \cdot \text{PD}_i + X_i'\Gamma + \alpha_{j(i)} + \delta_{n(i)} + \epsilon_{2i} \tag{2}$$

where $\delta_{n(i)}$ is a fixed effect for each position on the indictment (e.g., first, second) and $n(i)$ is a mapping from defendant i to his position on the indictment n. The identifying assumption is that once we condition on the defendant's position on the indictment there are no unobserved confounders that are correlated with both the defendant's potential outcomes (i.e., culpability) and the assignment to a PD. In Section (IV), I document that in federal district courts, unlike San Francisco, there is within-case selection that can confound a causal interpretation of model (1). However, after conditioning on the defendant's position on the indictment, balance tests on observable characteristics suggest that estimation via OLS recovers a causal relationship.

IV Identification strategy: Conflict-of-interest considerations in cases of multiple defendants

In multiple defendant cases, the public defender's office is usually constrained to represent only a single defendant due to potential conflicts of interest (Moore, 1984; Allison, 1976; Lowenthal, 1978; Prado et al., 1993). The Committee to Review the Criminal Justice Act, 1991–1992, determined that a "defender organization cannot properly undertake the representation of more than one defendant in a multi-defendant prosecution because a conflict of interest almost invariably results." The review committee specifically states that "private attorneys provide representation in multi-defendant and other cases in which representation by the federal defender could potentially create a conflict of interest."

In such circumstances, a PD is assigned to one of the indigent defendants, and the others are appointed to CAs. Figure (2) shows the average number of PD by the number of defendants in the case, and Figure (3) shows the distribution of defendants across attorney types by the number of defendants in the case. The figures clearly validate the conflict-of-interest hypothesis that the public defender organization will usually not represent more than one defendant within a multiple defendant case in both San Francisco and federal district courts.

The within-case comparison can be viewed as matching together similar units and then "randomly" flipping a coin to assign some to the treatment and the others to the control. Naturally, the setting, rather than statistical methods designed to optimize covariate balance, creates the matches. The matches are pre-determined outside of the control of the researcher. Thus, covariate balance can be used as a testable implication to validate the assumption that treatment was exogenously assigned within each case. The identifying assumptions in San Francisco and the federal system are different. In San Francisco, I argue that the within-case assignment to PD can be plausibly considered as good as random random. However, in federal courts it is necessary to control for the position of the defendant on the indictment to obtain exogenous treatment assignment.[13]

IV.A Overcoming selection in the assignment of defendants between PDs and CAs

I document extensive selection in the assignment of defendants between PDs and CAs, which is essential to overcome in order to understand whether PDs and CAs provide the same level of legal representation. If the cases that are assigned a PD are different in their severity and complexity as compared to those that are assigned a CA, then these differences need to be taken into account when the case outcomes are compared. To summarize the differences in the charges that defendants who are assigned a PD relative to a CA are facing, I use covariate indices that are based on a Oaxaca decomposition. In Appendix (D), I describe the exact construction of the covariate indices that are used both to document selection and to test for balance within a multiple defendant case. In both federal district courts and San Francisco, I observe offense codes that are highly predictive of the case outcomes but are to numerous to show comparisons for each category separately. The dimensional reduction that is conducted using the summary covariate indices allows me to present one summary measure that includes imbalances in demographics, charge severity measures and criminal history all at once. It is important to note that I observe detailed demographic and criminal history information only in San Francisco, but not in federal district courts; however, in both judicial systems detailed informations on the type and severity of the filled charges is observed.

IV.B San Francisco

Table (3) shows the overall differences in defendants' characteristics between PD and CA. The CA attorneys represent significantly more African-Americans, females, and defendants who are facing felony level charges. Those are the average characteristics of cases in which the PD office encounters a conflict of interest, and are *not* cases that the PD chooses not to represent. To

[13] As the assignment to a PD or a CA is usually not random, there is a concern that in cases involving numerous defendants (e.g., 30, 50, 100) the independence assumptions may be less plausible. Many defendants' cases may begin to inherit selection problems characteristic of the full sample since there may be differences in defendants' probability of assignment to a PD. To mitigate concerns of selection bias in multiple defendant cases, I limit the cases in our sample to have no more than 10 defendants. Relaxing this assumption to cases involving no more than 5 or 20 does not change the results of the paper. The constraint is binding only in federal courts. In San Francisco, the vast majority of cases are with co-defendants, two defendants within a case, and there is a small number of cases with more than four defendants. There are no cases with more than 10 defendants in San Francisco.

empirically test for differences between defendants who are assigned a PD compare to a CA, I use the following econometric model:

$$X_i = \beta \cdot PD_i + \alpha_{j(i)} + e_i, \tag{3}$$

where the β coefficient is the average difference in characteristic X_i across attorney types. When case fixed effects $\alpha_{j(i)}$ are not included, the β coefficient is exactly the difference in means, and when they are included it is the difference in means within a case. A cross-case comparison, when fixed effects are not included, can be sensitive to omitted-variable bias if there is selection in the type of cases that are assigned a PD relative to a CA.

Table (4), columns (1) and (2), shows clear evidence of selection in the assignment of defendants between PD and CA. The defendants who are assigned a CA are charged with more severe offenses and face a longer expected imprisonment time if convicted. This pattern of non-random sorting is the result of two factors. The first is that in San Francisco, the PD office handles the vast majority of cases, which are mostly not felony cases. Second, in cases with more severe charges there is a higher likelihood of a conflict of interests (e.g., between co-defendants), which leads to a higher proportion of defendants who are assigned to a CA among defendants facing felony-level charges.

In column (3), I restrict attention to multiple defendant cases with both PD and CA; however, I do not take into account variation in case-level characteristics (e.g., number of defendants). The differences between attorney types in columns (2) and (3) are based on a comparison of defendants across court-cases. In multiple defendant cases, a cross case comparison can provide a false impression as the number of CAs changes with the number of defendants in the case while the number of PDs is fixed at one. When the severity of the charges increases with the number of defendants, it is necessary to adjust for case fixed effects, i.e., conduct a within-case comparison, to obtain a reliable estimate of the differences in charge characteristics between defendants who are assigned a PD relative to a CA *within* a case.

Table (4), columns (4) and (6), show that *within* a multiple defendant case the treated and control units are comparable in demographic characteristics and charge severity measures. The adjusted differences in means using case level fixed effects, columns (4) and (6), shows that the differences between treated and control units are especially small relative to the baseline means of each measure (see Table (1)).

Figure (4) provides a visualized summary of the estimation results of Table (4). Each point on the figure is a t-statistic of the β coefficient in equation (3). The figure visualizes how the selection in the attorney type assignment goes away once the comparison is conducted within a case.

Figure (5) reports the results of a joint F-test for whether the controls are predictive of the attorney type assignment. The figure reports the observed value of the test statistics, the F-stat, and its likelihood under the null distribution of random assignment of defendants to attorney types. The null distribution was generated by a Monte-Carlo simulation with 1,000 random permutations of the PD assignment within a case, in multiple defendant cases, and across cases in single defendant cases. It is clear that in single defendant cases the assignment is not done at random; however, in multiple defendant cases there is no evidence of sorting within a case and we cannot reject the null hypothesis of random assignment.

Table (A.2) documents substantial within-case variation in observables. For example, in 33.8% of the cases there is at least one defendant with a prior arrest and one without. In 22.2% of the cases there is at least one black and one non-black defendant. The above analysis shows that this variation is *not* correlated with the within-case attorney type assignment, which supports the assumption that the multiple defendant scenario can be considered a natural experiment with exogenous PD assignment within a case.

IV.C Federal courts

The mechanism in which indigent federal defendants are appointed a legal counsel is different between districts and over time. Prado et al. (1993) describes the CA appointment process: "Some districts have systems in place to ensure an objective rotational system while others base an assignment decision on personal knowledge of an attorney's ability and skill level. In some districts the federal defender office assigns cases; in some districts an employee of the court is given the responsibility." A federal district has broad discretion in how it supplies indigent defense services. 18 U.S. Code § 3006A requires each federal district to prepare an indigent defense plan and to approve it by the judicial council of the circuit (see Chapter 2, § 210.10.10 (d), Appx 2A).[14] The indigent defense plan is obliged to satisfy a list of requirements, one of which is that "private attorneys shall be appointed in a substantial proportion of the cases" (18 U.S Code § 3006A(a)(3)); where "a substantial proportion" is interpreted as 25% of all indigent defense appointments on an annual base (Chapter 2, § 210.10.10 (d), Appx 2A).

Overall, defendants who are assigned to a PD organization are significantly different in the charge characteristics as compared to defendants who are assigned a CA. To empirically test for differences in observable characteristics between defendants assigned to a PD relative to a CA, I employ the following econometric models:

$$X_i = \beta \cdot PD_i + \gamma_{d(i)} + \eta_{t(i)} + \xi_{1i} \tag{4}$$

$$X_i = \beta \cdot PD_i + \alpha_{j(i)} + \xi_{2i} \tag{5}$$

$$X_i = \beta \cdot PD_i + \alpha_{j(i)} + \delta_{n(i)} + \xi_{3i}, \tag{6}$$

where γ_d, η_t, α_j, and δ_n are district, filing year, case and indictment order of appearance (e.g., first, second, third) fixed effects. These models are an analogue to model (3) for federal district courts,

[14] See this link for a template of a federal district indigent defense plan, http://www.uscourts.gov/file/vol07a-ch02-appx2apdf.

and the β coefficient can be interpreted as the difference in means in characteristic X_i between defendants who have been assigned a PD relative to a CA.

Table (5), columns (1) and (2), shows differences in the severity measures of the filed charges after adjusting for district and filed year fixed effects (i.e., model (4)). The comparisons are based on all indigent defendants (or all multiple defendants) and show that overall defendants who are assigned to a PD face less severe charges relative to defendants who are assigned a CA: a shorter predicted prison term, fewer felony-level charges, a slightly lower predicted probability of a conviction, and a lower predicted probability of a dismissal of charges. The results in columns (1) and (2) are similar to the ones in San Francisco and indicate similar pattern in both state and federal courts.

The main analysis sample is multiple defendant cases with both a PD and a CA within a case. This sample is different than the overall sample of indigent defendants in two respects: (i) It includes only individuals in multiple defendant cases, and (ii) It restricts attention to multiple defendant cases with both a PD and a CA. Table (5) shows that comparisons within multiple defendant cases suffer from the same selection patterns of the overall sample. For example, the cross-case comparison can compare a case with two CAs to one with a PD and a privately retained counsel.

Table (5), column (3), shows differences in charge severity between defendants assigned a PD and a CA within a case. This comparison reveals a reverse pattern of selection in the attorney type assignment compared to the cross-case comparison. The PDs are assigned to the defendant facing the more severe charges. In federal courts, the within-case comparison yields a bias estimate of the attorney type effect if the order of the defendants on the indictment is not taken into consideration, and column (3) documents this within-case selection pattern. Figure (6) shows the distribution of defendants across attorney types by the position of the defendant on the indictment in multiple defendant cases. This naïve comparison does not reveal differences in the probability of the first defendant on the indictment being assigned a PD relative to a CA. However, as the defendant is further down on the indictment his probability of being assigned a PD decreases dramatically. For example, among defendants who are listed on the first position of the indictment the share who are assigned a PD is approximately 40% relative to approximately 10% who are assigned a PD among defendants listed on the third position of the indictment. The position of the defendant on the indictment is a strong predictor of the attorney type that will be assigned to the defendant.

To examine whether the order of the defendants on the indictment is correlated with sentencing outcomes of interest such as length of incarceration I estimate the following model:

$$\text{asinh}(\text{Prison term})_i = \delta_{p(i)} + \alpha_{c(i)} + \kappa_i \tag{7}$$

As length of incarceration is an extremely skewed distribution a common practice is to perform some concave transformation (e.g., a logarithmic function). When the outcome of interest has a point mass at zero the asinh(·) function is commonly used as an approximation for the logarithmic function.[15] Figure (7) plots the estimated $\delta_{p(i)}$ coefficients and presents compelling evidence that the order on the indictment is a strong predictor of the length of incarceration that the defendant will be sentenced. The first defendant is likely to face a harsher sentence than the other defendants listed on the indictment. The order of defendants on the indictment can be considered as an additional measure of the differences in unobservable confounders (e.g., culpability) between the defendants.

To take into account the order of the defendants on the indictment, I estimates the model in equation (6) that includes a specific fixed effect for each position on the indictment. Table (5), column (4), shows that after conditioning on the defendant's order of appearance on the indictment there are *no* differences in the charge severity measures between defendants assigned a PD relative to a CA. Figure (8) provides a visualized summary of the balance tests reported in Table (5). The figure illustrates both the selection and non-random sorting that are present in a naïve comparison, and the comparability of defendants assigned to different attorney types within a case after conditioning on the position on the indictment.

As an additional balance test, I ranked the defendants within each case by their predicted prison term (months) and defined an indicator variable for the defendant who faced the highest predicted prison term within a case. In the same way I rank defendants within a case based on other predicted outcomes such as the probability of being convicted or the probability that a trial will take place. Table (6) reports the difference in means in the probability that the PD organization will be assigned to the highest-ranked defendant. Each row in the table refers to a ranking based on a different predicted sentencing outcome, and each cell is a coefficient from a different regression specification. The PD is more likely to be assigned to the more severe defendant within a case; however, this bias disappears once conditioning on the defendant's position on the indictment.

V Results

V.A San Francisco

The primary analysis sample is all indigent defendants in multiple defendant cases with both a PD and a CA. The main econometric model of interest is equation (1). Table (7) reports the estimation results. In the full sample with both single and multiple defendant cases, individuals who are first assigned a PD are sentenced to a shorter prison term by nearly 33.1% relative to those assigned a CA. This unadjusted difference falls to 18% with the inclusion of controls, which suggests that a naïve comparison can be influenced by selection bias in the assignment of defendants

[15]For example, see Gelber et al. (2016) who apply this approximation to Social Security Administration earning records or Card and DellaVigna (2013) who apply it to citations of academic papers, which is also a skewed distribution with a large mass at zero.

to different attorney types. Differences in *observable* defendant and charge characteristics explain a substantial share of the sentencing differences between those who are assigned a PD vs. a CA. Altonji et al. (2005) show that differences between the covariate-adjusted and unadjusted estimates are a measure of selection due to omitted variables, which re-enforces the claim that a simple regression that relies upon a strong, unverifiable conditional independence assumption will not identify a causal relationship.

Table (7), columns (4)–(6), shows that within multiple defendant cases, those assigned a PD are sentenced to a 10.5% *shorter* prison term relative to their co-defendants who are represented by a CA. The estimate with covariate adjustment (a 10.7% shorter prison term) is not statistically different from the unadjusted estimate, which stands in contrast to the differences in estimates with and without covariate adjustments in the full sample that includes single defendant cases. Column (6) reports the results (10.1%) once controlling for prior representation by a PD, which also does not impact the estimate. Assignment to a PD also decreases the probability of conviction by 3.9pp and any prison time by 1.8pp which is a 22% decrease relative to the mean rate of imprisonment. I find the attorney type of the defendant has no statistically significant effect on the sentenced jail term or the probability of being released on bail.

To understand the magnitude of the estimated effects it is necessary to compare the coefficients to the average values of the sentencing outcomes. Figure (9) summarizes the magnitude of the estimated attorney type effects within multiple defendant cases relative to the baseline mean of each sentencing outcome. The right plot presents confidence intervals for the estimated effects. The left plot illustrates the likelihood of the observed estimated effects relative to a null distribution in which the attorney type assignment has no effect. The null distribution was generated by a Monte-Carlo simulation with 1,000 random permutations of the PD assignment within a case. The black dots indicate the values of the coefficient that were obtained by a random permutation and darker areas represent values of the coefficient that are likely to be observed under random chance when the attorney type has no effect. The red triangles indicate the observed values of β in the data. The results suggest that overall PDs obtain significantly more favorable case outcomes for their clients in a range of sentencing outcomes. The largest effects are on the defendant's prison term, the probability of being sent to state prison at all, and the probability of conviction. The estimated effect on the probability of being sent for any period of time to a state prison is a 22% decrease relative to the baseline mean.

Next I investigate whether the attorney type assignment has heterogeneous effects across defendants, by interacting PD_i with various defendant characteristics:

$$Y_i = \beta \cdot \mathrm{PD}_i + \gamma \cdot \mathrm{PD}_i \times C_i + X_i'\Gamma + \alpha_{j(i)} + \epsilon_{3i}, \tag{8}$$

where C_i is one of the elements of X_i'.[16]

[16]The above specification includes all the main effects of the interaction term.

Table (8) reports the estimated γ coefficients for interactions on demographic, charge severity measures, and criminal history. The coefficients on the interactions are large in magnitude, yet imprecise and not statistically significant. To gain additional accuracy I use permutation inferences and examine the likelihood of observing the estimated coefficient under a scenario in which the attorney type assignment has no impact. In Appendix (E), I compare permutation inference and the usual cluster-robust standard errors in terms of power in my context and find that permutation inference can have substantially higher power to reject the null when it is false.[17]

Figure (10) reports the results of the heterogeneity analysis for three different outcomes. The figure shows that defendants who are facing more severe charge (felony vs. misdemeanor) and those who have a longer criminal history are the ones who stand to benefit the most from being represented by a PD relative to a CA. Although, the coefficients in Table (8) are imprecise, the results in Figure (10) clearly illustrates that the interactions of defendant charge severity and criminal history with attorney type are unlikely under random chance and are statistically significant when permutation inference is used. The heterogeneous effects are studied with respect to the probability of being imprisoned, the length of imprisonment, and the probability of conviction. Table (9) reports the P-values of the observed values of γ relative to the distribution under random change in Figure (10).

Providing defendants with a higher quality of legal representation can lead to fewer defendants being sent to prison, which might cause an increase in crime as they will not be incapacitated (Ater et al., 2017). I estimate equation (1), where the outcome, Y_i, is recidivism within a certain period of time (e.g., 10 weeks) from the date of disposition. Figure (11) reports the estimation results of β, which is not statistically significant, but is positive and increases until 60 weeks, at which point the coefficient stabilizes, and at around 120 weeks it starts to decline toward zero.[18]

From a constitutional perspective, if defendants in the same case have been assigned attorneys with varying levels of legal expertise to such a degree that it influenced their sentencing outcomes, then there is a concern about a violation of the defendants' Sixth Amendment rights.

V.B Federal district courts

In federal courts, to estimate the attorney type causal effect it is also necessary to condition on the order of appearance of the defendants on the indictment (i.e., model (2)). The estimation results are for cases that have been terminated between 1996 and 2014 in federal district courts. Table (10) shows the estimation results and it highlights the selection challenges that must be overcome to estimate the causal effect of attorney type on defendant's sentencing outcomes. The unadjusted

[17] This is *not* a general result that can be applied to other cases, but rather specific to my application. In Appendix (E) I use Monte-Carlo simulations to compare the power of the two methods of conducting inference. All the details are described in the appendix.

[18] Measuring recidivism from the time of disposition captures incapacitation effects as well as any behavioral effects.

asinh(Prison term) estimate in column (1), -0.278, is similar to the estimate in San Francisco, -35.7, and after adjusting for the charge codes the coefficient shrinks to -0.0322. Unlike San Francisco, in federal courts the order of the defendants on the indictment has a large impact on the estimates and a simple within-case comparison yields a coefficient of 0.115 after covariate adjustment. The within-case comparison in federal courts yields an opposite result to the estimates in San Francisco. However, after controlling for the defendant's position on the indictment (columns (7) and (8)), the estimated effect, -0.0462, has the same sign as the one in San Francisco, -0.109.

According to Table (10), PDs obtain shorter prison sentences for their clients (4.64%), a slightly higher probation term (2.39%), and a lower probability of any prison term (0.819pp). I find no differences in the probability of reaching a plea bargain on some of the charges; however, I find a 0.768pp statistically significant lower probability of taking a case to trial. This is a small estimated coefficient, but relative to the average number of cases that go to trial (5% in this sample) it implies a 15.6% decrease in the probability that a trial will take place. I find no differences in the probability of conviction or acquittal of defendants initially assigned a PD relative to a CA.

Taken together, the findings from both San Francisco and federal district court present empirical evidence that the method by which indigent defense services are provided, PD vs. CA, influences the trial outcomes of the defendants. Indigent defendants in multiple defendant cases who were assigned a PD obtained more favorable outcomes than their co-defendants who were represented by a CA.

V.C Attorney characteristics

The differences in case outcomes that have been documented above can be the result of several mechanisms. First, attorneys who select to work in a PD office can have different characteristics (e.g., experience) than those who work as private attorneys and accept appointments from the court (CA). Second, a defendant assigned to a PD office is represented by an organization and not only by a specific attorney. Within the PD office the attorney that is assigned to represent the defendant can consult with other professionals in the office and be exposed to organizational norms and knowledge that have been accumulated through past representation of similar cases.

In the court records of San Francisco, I observe the name of the attorney that represented the defendant and his type (e.g., CA, PD). I use name tabulations from the US Census 2000 and the Social Security Administration to infer the race, ethnicity, and gender of the attorneys from their names. Information on the institutions that awarded B.A. and J.D. degrees to the attorney was obtained using the search engine of the state bar association. To obtain the ranking of each institution I use the information that is publicly available on U.S. News.[19]

Table (11) documents the characteristics of PDs and CAs. Relative to CAs, PDs are younger,

[19] U.S. News publishes a ranking of universities and colleges in the U.S. The ranking can be for the entire institution or for a specific program such as a law school. https://www.usnews.com/.

less experienced, demographically more diverse, and studied in more selective colleges (the best ranking for a university/college is number 1). Two factors that can explain why young individuals who obtained their J.D. in high-ranked universities choose to work in a PD office are (i) ideological motivation and (ii) financial incentives. Regarding the ideological motivation, PDs may desire to represent individuals who cannot afford to hire a private attorney, and will be over-represented by individuals from minority communities compare to the general population. As for financial, Field (2009) documents how in recent years higher ranked J.D. programs provide fee remissions and subsidies to students who work in public interest jobs after graduation. Working in a PD office is considered a public-interest job, unlike being a CA.

To understand how much of the causal attorney type effect can be explained by attorney characteristics, I add them as additional controls to the regression specification in equation (1). Attorney characteristics account for slightly less than half of the estimated attorney type effects on the sentenced prison length, whether the defendant serves time in prison, and his probability of conviction (see Table (12)). Once attorney characteristics are controlled for the estimated differences in sentencing outcomes have the same sign, but become smaller in magnitude, and not statistically significant.[20]

In addition, representation by an organization is inherently different than being assigned to a specific attorney. Indigent defendants are not assigned to an individual attorney in the PD office by the court, but rather to an organization. The PD office determines how to divide the workload among its attorneys. One attorney can represent the defendant at the initial stages of the case and another at the more advanced court proceedings, including the plea negotiations. Table (13) reports the differences in the characteristics of the attorney that first represents the defendant and the terminating attorney. Overall, defendants assigned to the PD office change individual attorneys more often, 52%, relative to 20.7% among those assigned a CA. It is also more frequent that the terminating attorney has more years of experience than the initial attorney among defendants assigned to the PD office.

VI Discussion

The vast majority of defendants facing criminal charges require the assistance of court-appointed legal counsel. This paper develops a framework to compare two methods of providing legal representation to defendants who cannot afford to hire an attorney in the private market. I use a new empirical identification strategy and administrative court records from both state and federal district courts to study whether defendants assigned to a public defender (PD) obtain better or worse

[20]These results relate to the findings in Abrams and Yoon (2007), who utilized the quasi-random assignment of attorneys to cases within the public defender's office in Nevada, to estimate the impact of attorney characteristics (e.g., experience, gender, race) on case outcomes. They find that attorney characteristics have statistically significant prediction power on the sentencing outcomes of the case.

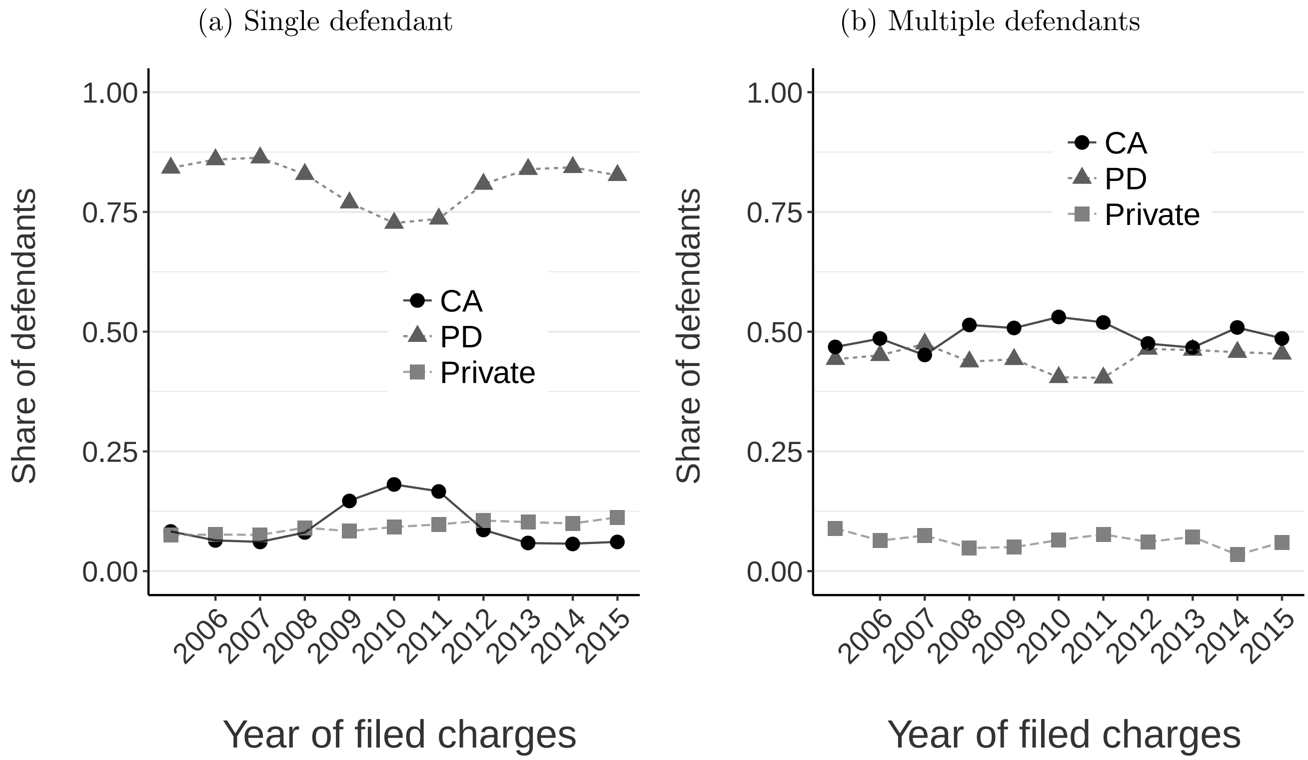
case outcomes than those who are represented by a private court-appointed attorney (CA). The results have direct policy implications on how indigent defense representation should be provided and whether the current system violates defendants' Sixth Amendment rights.

Defendants who have been represented by a PD, relative to a CA, obtained more favorable case outcomes (e.g., shorter prison sentences, lower probability of any imprisonment). In San Francisco, defendants who face more severe charges (felony vs. misdemeanor) and have a longer criminal history are the ones driving the results. Those who face a higher risk of imprisonment are the ones on whom the attorney type makes the largest impact. One explanation for these differences is that attorneys who work for a PD organization are substantially different in their observable characteristics from those who self-select to act as CAs. PDs have fewer years of experience, are demographically more diverse, and studied in more selective institutions both in their B.A. and J.D. programs.

The method of provision of indigent defense services is part of the bigger question of how the state should supply public services (e.g., police, prison). Should the state establish a PD organization or use the private sector and hire CAs? To answer this question, one of the key issues that needs to be addressed is which kind of attorney will select to represent low-income defendants under each one of the aforementioned alternatives. Future research is needed to examine what motivates attorneys to select to work in a PD office, relative to the self-selection of those who act as CAs. More information is needed to understand how policy makers can mitigate the attorney type differences that are documented in this study.

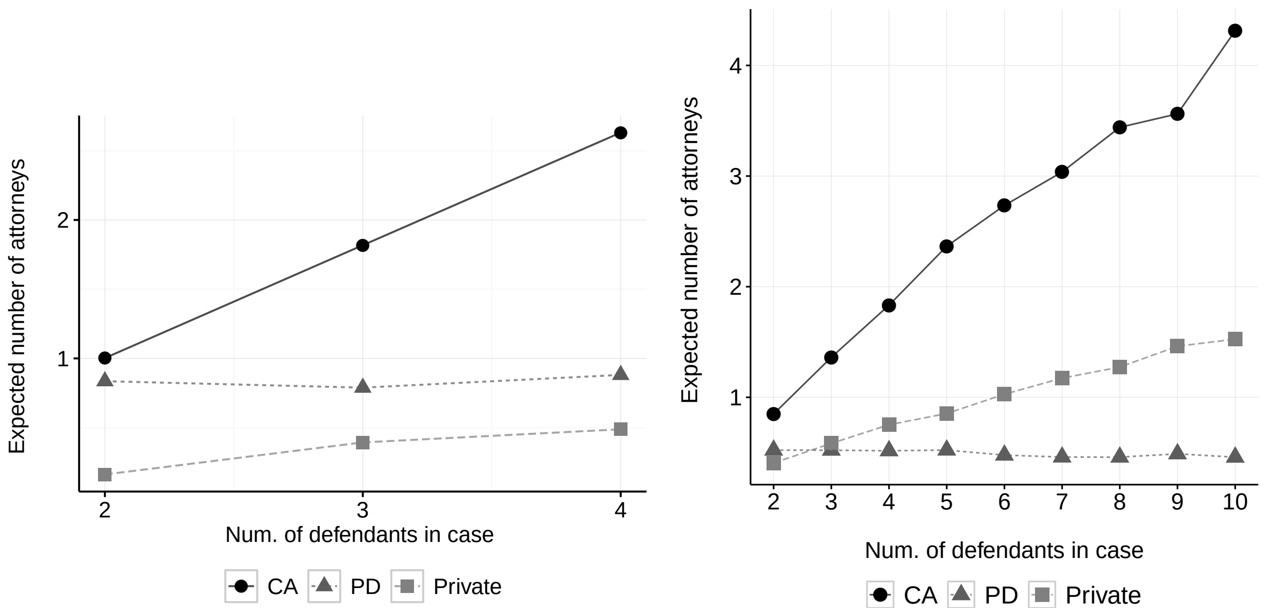
Figures

Figure 1: San Francisco: The distribution of defendants across attorney types and over time, by filing year



Notes: The figure shows the distribution of criminal defendants in San Francisco across attorney types. The left plot shows the distribution of defendants across attorney types in cases with a single defendant. The right plot shows the distribution of defendants across attorney types in cases with multiple defendants.

Figure 2: Validating the conflict of interest hypothesis

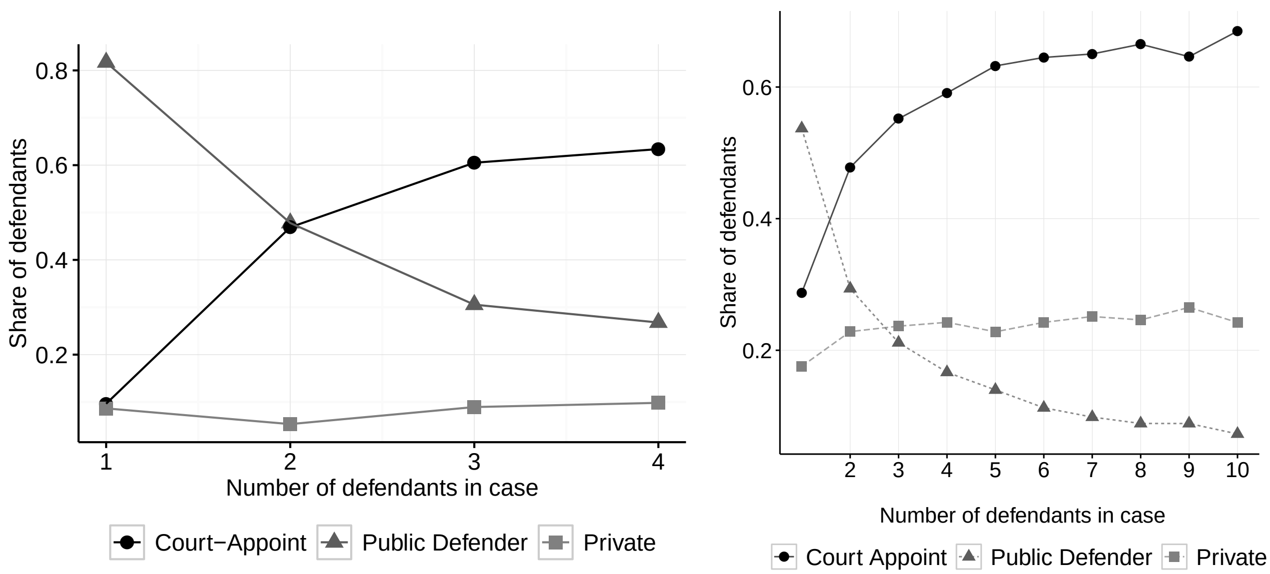


(a) San Francisco

(b) Federal district courts (1996 − 2014)

Notes: The figure presents descriptive evidence that validate the conflict of interest hypothesis that a PD organization cannot usually represent more than one defendant within a multiple defendant case. The x-axis describes the number of defendants in a case and the y-axis the average number of attorneys from each type (e.g. PD or CA). For example, Panel (a) shows that in multiple defendant cases in San Francisco with three defendants there are on average almost 2 CAs and approximately one PD. Panel (b) shows the evidence for cases terminated in federal district courts from 1996 to 2014.

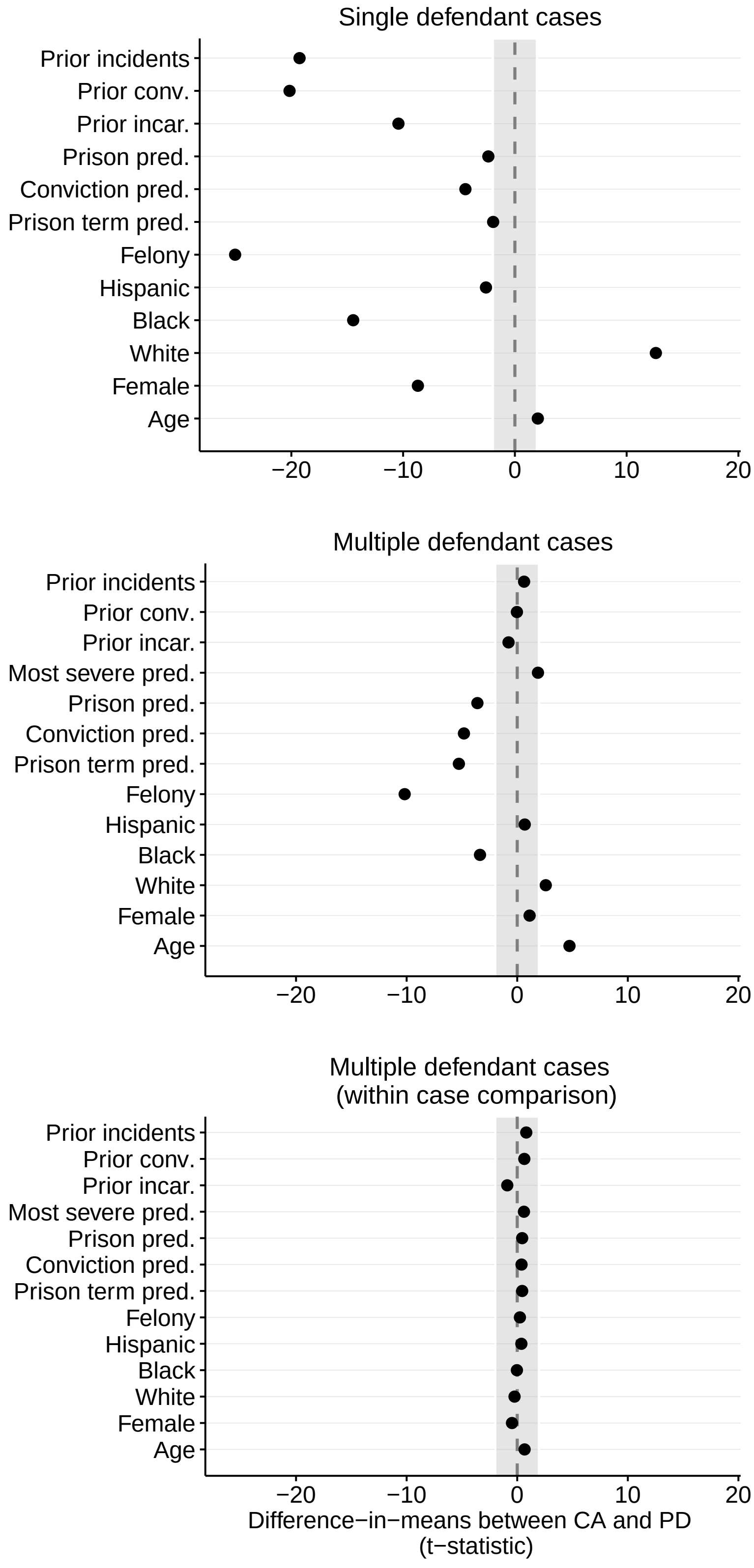
Figure 3: Distribution of defendants across attorney types, by the num. of defendants in the case



(a) San Francisco (b) Federal district courts (1996 – 2014)

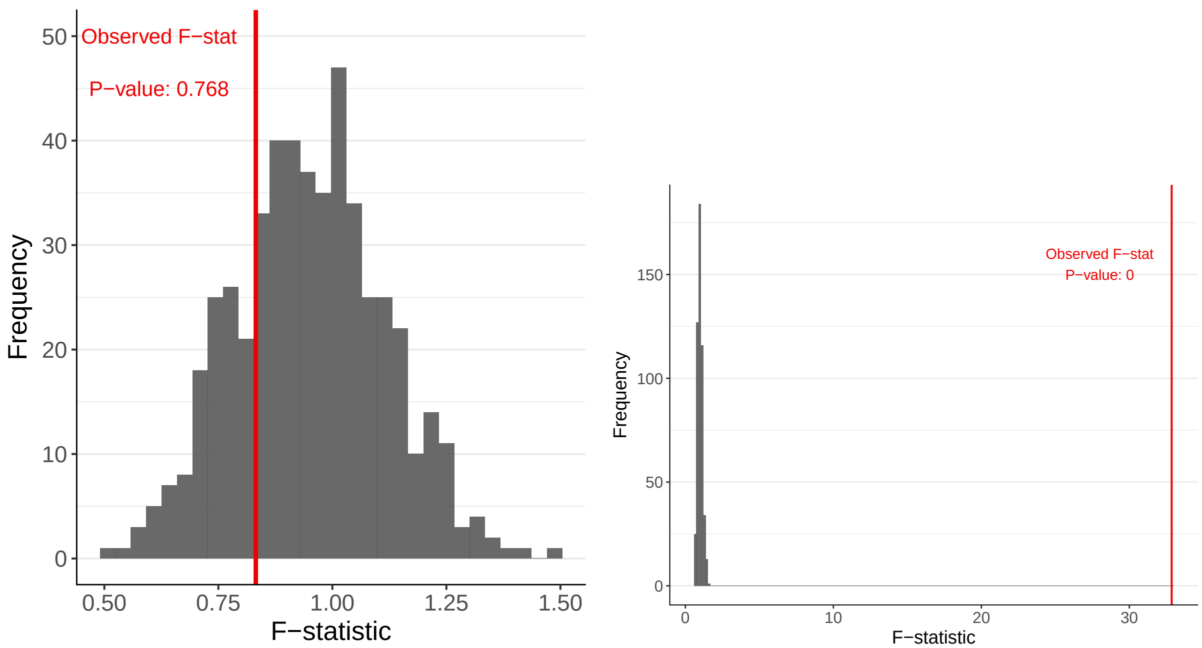
Notes: The figure shows the distribution of attorney types by the size of the multiple defendant case. As the number of co-defendants in a case increases the share who are assigned to a PD decreases and the share who are assigned a CA or represented by a private attorney increases. This pattern is another validation of the conflict of interest hypothesis.

Figure 4: San Francisco: Differences in observable characteristics between defendants assigned to PD vs. CA $(2006 - 2016)$



Notes: Each point on the figure is a t-statistic of the β coefficient in equation (3). Standard errors are clustered at the case level. The two upper plots show the results from specifications without case FE, and the bottom plot reports the results when case FEs are included. The gray area represents the confidence interval in which the null that the coefficient β is zero cannot be rejected. Since the number of observations when estimating a given specification for each of the outcomes is the same; and the figure reports t-statistics, rather than β coefficients, then the gray area is the same for all the t-statistics of each of the outcomes and is approximately ± 1.964 around zero.

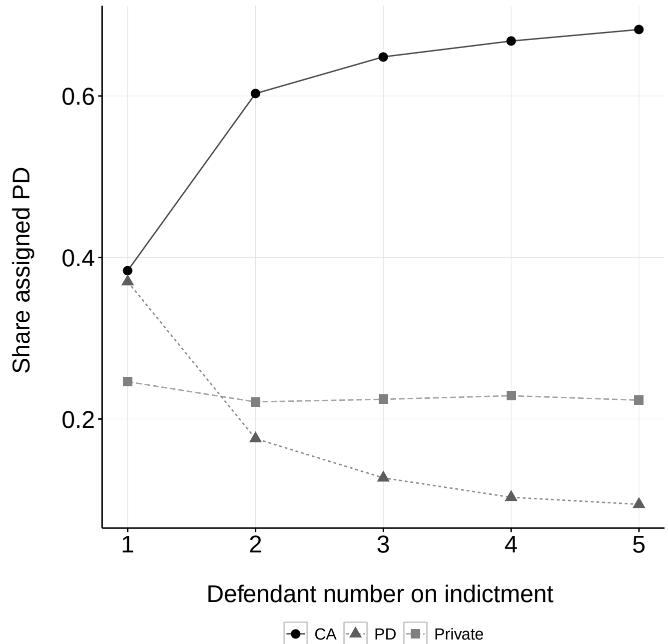
Figure 5: San Francisco: Monte-Carlo permutations of attorney type assignment within a case: F-statistic using Offense codes and controls



(a) Within multiple defendant case (b) Across single defendant cases

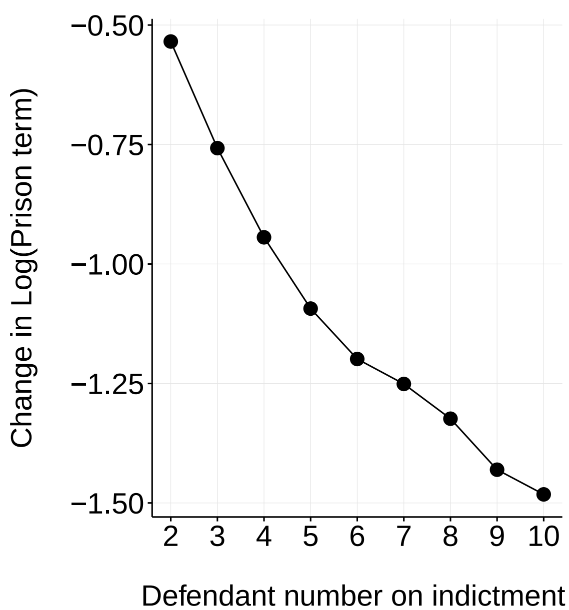
Notes: Each one of the plots above uses Monte-Carlo simulations to asses whether the observed F-statistic is likely under a mechanism which randomly assigned defendants across attorney types. Fischman (2011), Abrams et al. (2012) and Abrams and Fackler (2017) all used similar Monte-Carlo simulation procedures when assessing covariate balance and to correct finite sample coverage concerns with the asymptotic distribution of the conventional F-statistic. The red line shows the observed F-statistic and the histogram plots an approximation of the distribution of the F-statistic under a random assignment mechanism using 1,000 random re-labellings of defendants across attorney types. I randomly permuted/shuffled which defendants have been assigned to a PD relative to a CA, and then estimated the F-statistic for the null that all the coefficients are equal to zero. In the multiple defendant sample the permutations are done within a case. The number of re-labellings we use is 1,000 and it is similar to what is commonly used in the statistics literature. For example, Athey et al. (2018) and Anderson and Magruder (2017) use 1,000 draws/re-labellings; and Keele and Miratrix (2017) use 500 draws/re-labellings.

Figure 6: Order on the indictment and the probability to be assigned PD among multiple defendant cases



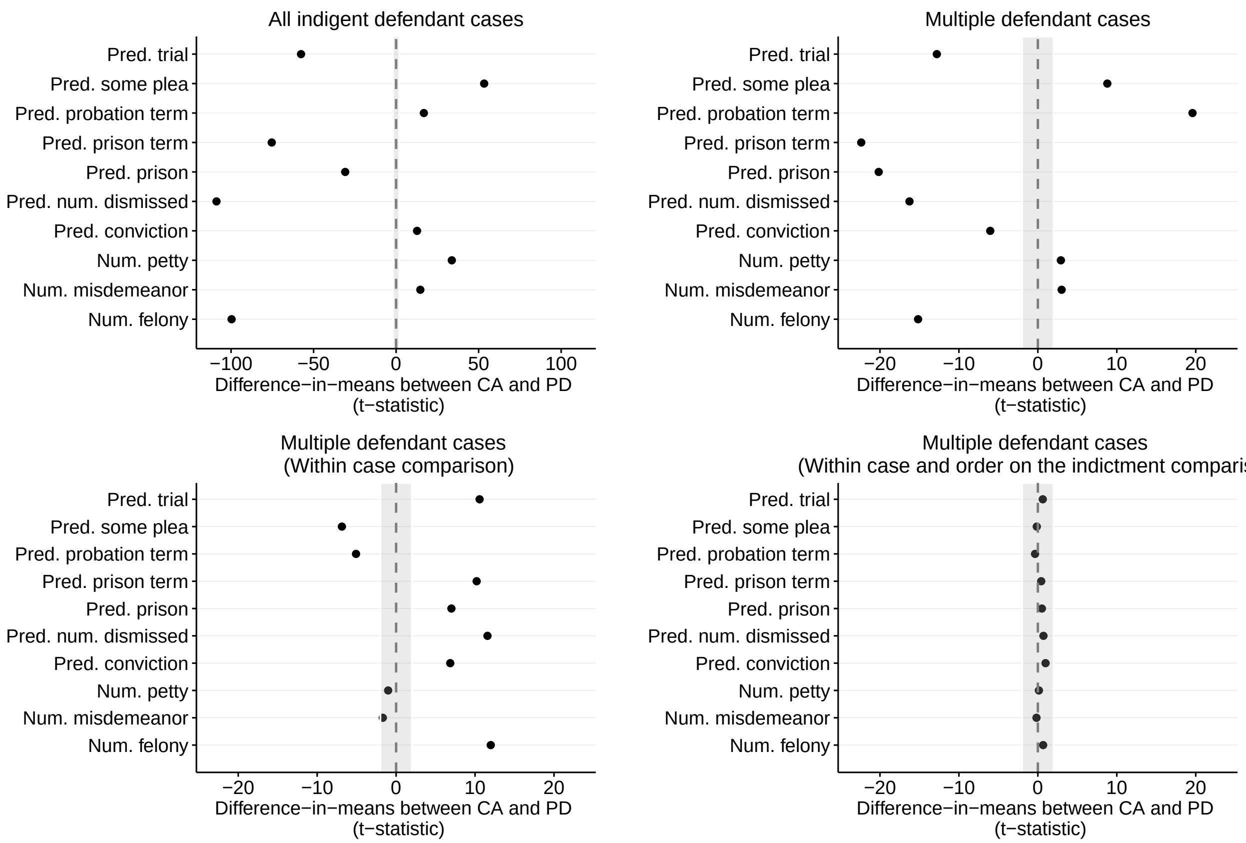
Notes: The figure reports the distribution of defendants across attorney types by the number of the defendant on the indictment. For example, among defendants who are listed first on the indictment approximately 40% will be assigned to a PD, another 40% a CA, and the remaining 20% will be represented by a private attorney. The share of defendants who are represented by a PD is decreasing by the number of the defendant on the indictment.

Figure 7: Order on the indictment and the length of incarceration



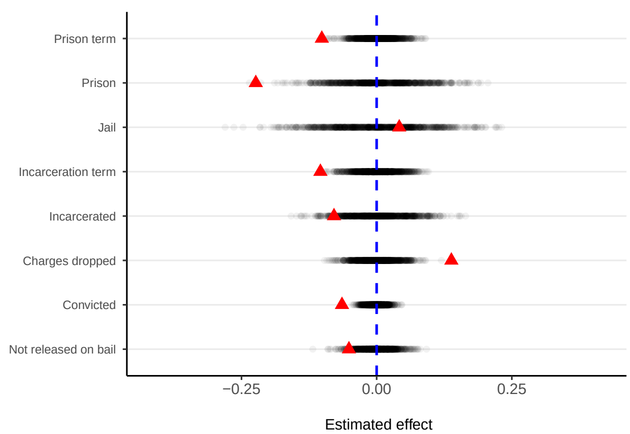
Notes: The figure reports the δ coefficients from equation (7). Each point is a fixed effect for a different position on the indictment, where the omitted category is the first defendant and all the coefficients report relative differences compare to the defendant that is listed first on the indictment.

Figure 8: Federal courts: Differences in observable characteristics between defendants assigned to PD vs. CA (1996 – 2014)

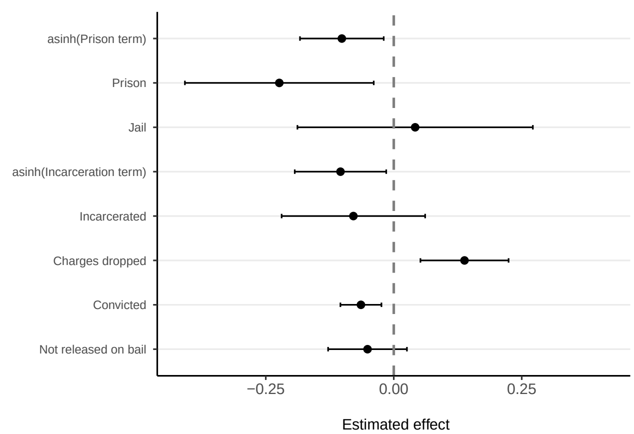


Notes: Each point on the figure is a t-statistic of the β coefficient from model specification (4), or (5) or (6). Standard errors are clustered at the case level. The two upper plots show the results from specifications without case FE, and the bottom figures reports the results once including case FE. The bottom right (left) figure includes (does not include) controls for the defendant's position on the indictment. The standard errors are cluster-robust at the case level. The gray area represents the confidence interval in which the null that the coefficient β is zero cannot be rejected. Since the number of observations when estimating a given specification for each of the outcomes is the same; and the figure reports t-statistics, rather than β coefficients, then the gray area is the same for all the t-statistics of each of the outcomes and is approximately ± 1.964 around zero.

Figure 9: San Francisco: The effects of the attorney type on the defendant's case outcomes



(a) Permutation inference

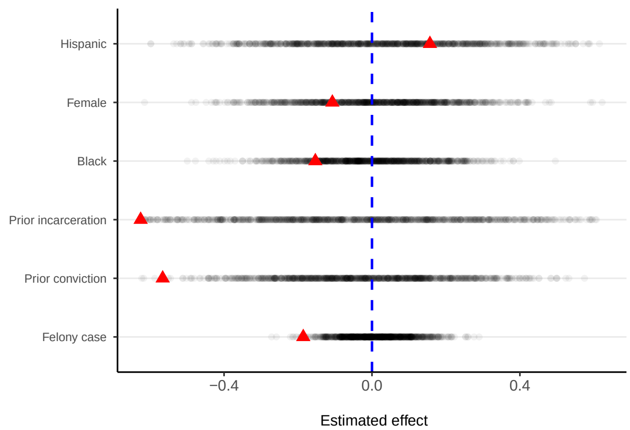


(b) Cluster-robust standard errors

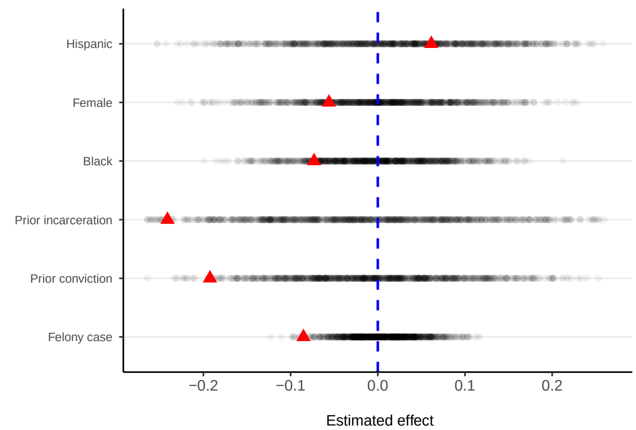
Notes: Confidence intervals for selected coefficients from Table (7) are presented divided by the baseline mean value of the sentencing outcome of interest. As in Table 7, standard errors are clustered at the case level. The right plot displays 1,000 estimates of the β coefficient from equation (1), where defendants in the same case have been randomly labelled as being represented by a PD—the randomization is within a case. The black dots are from each one of the permutations. Areas which are darker indicate a higher chance of observing those values due to random chance, when the attorney type has no impact on the case outcomes. The red triangle indicate the observed values in the data, which are not likely under the null that the attorney type has no effect on the case outcomes. The number of re-labellings we use is 1,000 and it is similar to what is commonly used in the statistics literature. For example, Athey et al. (2018) and Anderson and Magruder (2017) use 1,000 draws/re-labellings; and Keele and Miratrix (2017) use 500 draws/re-labellings. The Log($\cdot$) function is approximated using the arcsinh function which is commonly used as an approximation to the logarithmic function when the variable of interest contains values of zero.

Figure 10: San Francisco: Heterogeneity in the effects of the attorney type on the defendant's case outcomes by criminal history and demographic characteristics

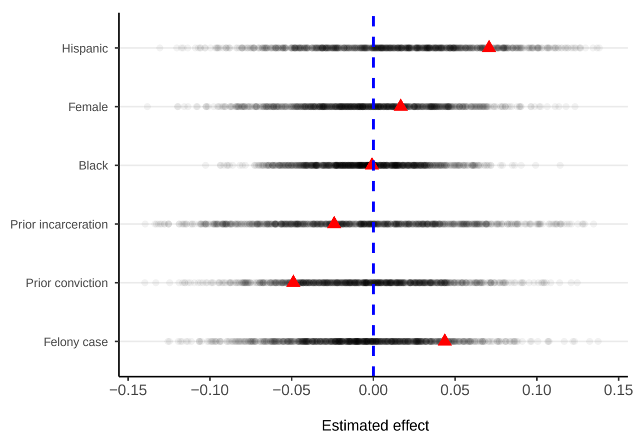
(a) Any prison



(b) Prison term (asinh)

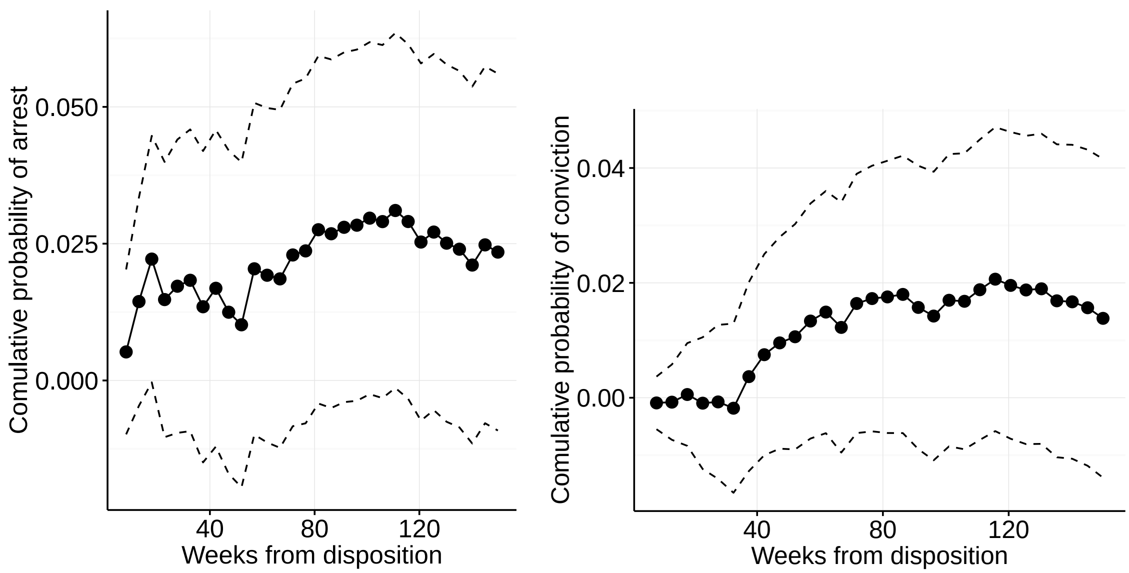


(c) Conviction



Notes: See the notes to figure (9). The only difference is that here the estimates are not of the *beta* coefficient from equation (1), but rather the γ coefficient from equation (8) for different dimensions of possible heterogeneity (e.g., gender, criminal history). Each one of the plots is for a different sentencing outcome (e.g., any prison sentence, conviction).

Figure 11: San Francisco: The relationship between initial assignment to a PD and involvement in the criminal justice system within a fixed period of time since disposition



(a) New arrest with X weeks from disposition

(b) New conviction with X weeks from disposition

Notes: Each point in the figures is the estimated β coefficient from equation (1), where the outcome, Y_i, is a new arrest/conviction within a certain period of time from the date of disposition. The x-axis measures the time from disposition in weeks. The standard errors are cluster-robust at the case level. The recidivism measures are calculated only using new offenses/convictions in San Francisco.

Tables

Table 1: San Francisco, defendants' characteristics in single and multiple defendant cases

	Single defendant (1)	Multiple defendant (2)	Multiple design (3)
Age	35.503	32.250	32.181
White	0.530	0.453	0.435
Black	0.440	0.515	0.535
Female	0.177	0.239	0.238
Hispanic	0.195	0.200	0.201
Highest filed charge felony	0.518	0.829	0.873
Predicted conviction	0.586	0.614	0.619
Predicted prison>0	0.057	0.058	0.058
Num. prior incarcerations	0.174	0.146	0.158
Num. prior convictions	0.521	0.470	0.485
Num. prior arrests	2.237	2.027	2.113
Dropped charges	0.246	0.249	0.230
Convicted	0.598	0.586	0.609
Prison	0.063	0.080	0.079
Jail	0.098	0.070	0.071
Observations	64,191	9,576	7,164

Notes: The table presents descriptive statistics for all criminal defendants in San Francisco between 2006 and 2016. Columns (1) and (2) include all incident-defendant pairs in the analysis data set and are not restricted to indigent defendants. Columns (3) includes only indigent defendants in multiple defendant cases which have both a PD and a CA. The third columns reports the descriptive statistics for the main analysis sample in which the assignment of defendants to attorney type is as good as random within a case.

Table 2: Descriptive statistics on indigent defendants in federal courts (1996 − 2014)

	All	Single def.	Multiple def.	Multiple def. (PD & CA)
	(1)	(2)	(3)	(4)
Prison term (months)	38.970	30.360	61.070	54.730
Prison	0.860	0.860	0.850	0.860
Conviction	0.950	0.960	0.940	0.940
Some plea	0.930	0.940	0.890	0.900
Trial	0.030	0.020	0.060	0.050
Acquittal	0	0	0.010	0.010
Predicted prison term	40.640	32.320	61.980	57.710
Predicted prison	0.850	0.850	0.860	0.870
Probation term	3.100	3.120	3.050	2.960
Predicted plea	0.920	0.930	0.900	0.900
Predicted trial	0.040	0.030	0.060	0.050
Predicted conviction	0.950	0.950	0.940	0.940
Predicted num. convictions	1.120	1.070	1.260	1.230
Predicted num. dismissed	0.710	0.430	1.420	1.360
Num. felony	1.680	1.310	2.610	2.520
Number of defendants	2.350	1	5.820	5.480
CA in case	0.510	0.350	0.930	1
PD in case	0.630	0.650	0.590	1
Observations	651,666	468,791	182,875	84,260

Notes: The table presents descriptive statistics for all criminal defendants in cases terminated in federal district courts from 1996 to 2014. The four columns refer to different sub-samples of the data. The column (1) makes no restrictions and includes all defendants. Column (2) restrict attention to individuals in single defendant cases, without any co-defendants. Column (3) Restrict the sample to individuals in cases that includes more than one defendant. Column (4) restrict the sample in column (3) to multiple defendant cases in which at least one defendant is represented by a PD and another by a CA. In each of these cases there are both defendants who are represented by a PD and a CA, which allows to conduct a within case comparison of attorney types. All the predicted variables are summary measures for the charges that have been filed against the defendant based on a Oaxaca decomposition. Appendix (D) describes how the predicted variables are constructed. Table (A.1) in Online Appendix (A) presents similar descriptive information for cases that have been terminated in federal district courts between 1970 to 1995.

Table 3: San Francisco: Differences in defendants' characteristics between PD and CA across all cases

	Ave. CA	Ave. PD	P-value of difference
Age	35.290	33.666	0.000
Female	0.459	0.550	0.000
White	0.176	0.226	0.000
Black	0.515	0.431	0.000
Hispanic	0.194	0.204	0.021
Highest filed charge felony	0.543	0.763	0.000
Predicted Prison term	11.038	11.808	0.000
Predicted conviction	0.585	0.603	0.000
Predicted prison>0	0.057	0.059	0.000
Num. prior incarcerations	0.175	0.228	0.000
Num. prior convictions	0.512	0.767	0.000
Num. prior arrests	2.236	3.109	0.000

Notes: Each cell in the table reports the average value of each of the covariates. The predicted sentencing outcomes are summary measures that include indicators for broad charge categories (e.g. robbery, larceny, drug). Online Appendix (D) describes how the predicted summary measures are constructed.

Table 4: San Francisco: Differences in observable characteristics between defendants who are assigned PD and CA

	All indigent	All Multiple	Multiple PD & CA		Co-defendant PD & CA	
	(1)	(2)	(3)	(4)	(5)	(6)
Age	1.624***	1.064***	0.630**	0.158	0.225	0.225
	(0.124)	(0.252)	(0.280)	(0.292)	(0.312)	(0.337)
Female	−0.050***	0.010	−0.002	−0.006	−0.007	−0.007
	(0.004)	(0.009)	(0.010)	(0.013)	(0.011)	(0.015)
White	0.084***	0.022**	0.004	0.002	−0.003	−0.003
	(0.005)	(0.010)	(0.012)	(0.011)	(0.013)	(0.013)
Black	−0.091***	−0.029***	−0.004	−0.006	−0.001	−0.001
	(0.005)	(0.011)	(0.012)	(0.011)	(0.013)	(0.012)
Hispanic	−0.010**	0.005	0.002	0.001	0.007	0.007
	(0.004)	(0.008)	(0.009)	(0.013)	(0.011)	(0.015)
Felony	−0.220***	−0.069***	−0.007	0.001	0.001	0.001
	(0.005)	(0.008)	(0.008)	(0.003)	(0.009)	(0.003)
Predicted prison term	−0.771***	−0.783***	−0.170	0.008	0.089	0.089
	(0.090)	(0.162)	(0.178)	(0.180)	(0.175)	(0.203)
Predicted convicted	−0.018***	−0.009***	−0.0001	0.001	0.001	0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Predicted prison	−0.002***	−0.002***	−0.00000	0.0001	0.0004	0.0004
	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Most severe	–	0.023**	0.025**	0.008	0.015	0.015
	(0.003)	(0.011)	(0.012)	(0.022)	(0.013)	(0.026)
Num. prior incarceration	−0.053***	−0.008	−0.001	−0.010	−0.018	−0.018
	(0.007)	(0.011)	(0.013)	(0.017)	(0.015)	(0.019)
Num. prior convictions	−0.256***	−0.001	0.056**	0.037	0.025	0.025
	(0.015)	(0.025)	(0.028)	(0.034)	(0.032)	(0.039)
Num. prior incidents	−0.874***	0.053	0.193*	0.101	0.115	0.115
	(0.054)	(0.093)	(0.104)	(0.119)	(0.121)	(0.141)
Observations	67,620	8,975	7,164	7,164	5,826	5,826
Case FE	No	No	No	Yes	No	Yes

Notes: Each cell in the table contains the coefficient on an indicator whether the defendant was initially assigned a PD. The table reports the estimates of the β coefficient from model (3). Standard errors are clustered-robust at the case level. Columns 3 and 4 include all multiple defendant cases with both a PD and a CA within each case. Columns 5-6 include only multiple defendant cases with exactly two indigent defendants that one was assigned a PD and the other a CA. For example, a case with 3 indigent defendants that two of which are represented by CAs and the third by a PD will be included in columns 3 and 4 but not in columns 5 and 6. Notice also that in columns 5 and 6 the number of individuals within each case that are assigned to a PD is exactly the same as the number that is assigned to a CA. In this type of a balanced design the estimates in columns 5 and 6 are mechanically the same; however, the standard-errors are affected by the inclusion of case-level FEs in the regression specification. This mechanical equality between columns 5 and 6 in the point estimates would not have hold if continuous control variables would have also been included in the right hand side of the regression specification.

*p<0.1; **p<0.05; ***p<0.01

Table 5: Federal courts, terminated cases 1996 − 2014: Difference in filed charges between defendants assigned to PD and CA across samples

	All indigent	All multiple	Multiple (PD & CA)	Multiple (PD & CA)
Num. felony	-0.450***	-0.121***	0.0688***	0.00470
	(0.00451)	(0.00801)	(0.00572)	(0.00648)
Num. misdemeanor	0.0142***	0.00378**	-0.00124	-0.000147
	(0.000981)	(0.00126)	(0.000737)	(0.000856)
Num. petty	0.0180***	0.00158**	-0.000487	0.0000695
	(0.000536)	(0.000541)	(0.000483)	(0.000524)
Predicted prison term	-9.086***	-4.809***	1.212***	0.0613
	(0.121)	(0.215)	(0.119)	(0.135)
Predicted prison	-0.0117***	-0.0114***	0.00192***	0.000173
	(0.000380)	(0.000566)	(0.000273)	(0.000320)
Predicted probation term	0.168***	0.341***	-0.0380***	-0.00321
	(0.0100)	(0.0174)	(0.00747)	(0.00887)
Predicted some plea	0.00697***	0.00195***	-0.000909***	-0.0000194
	(0.000130)	(0.000222)	(0.000133)	(0.000153)
Predicted trial	-0.00698***	-0.00294***	0.00163***	0.000114
	(0.000121)	(0.000230)	(0.000154)	(0.000174)
Predicted conviction	0.00105***	-0.000713***	0.000550***	0.0000968
	(0.0000811)	(0.000120)	(0.0000798)	(0.0000928)
Predicted mum. dismissed	-0.356***	-0.0941***	0.0478***	0.00356
	(0.00327)	(0.00581)	(0.00412)	(0.00468)
N	468,791	182,875	84,260	84,260
Def. Num. FE	No	No	No	Yes
Case FE	No	No	Yes	Yes
District FE	Yes	Yes	No	No
Year FE	Yes	Yes	No	No

Notes: Each cell in the table reports the coefficient of an indicator whether the defendant was initially assigned a PD or a CA. This is the β coefficient from estimating one of the models (4), (5), or (6). Standard errors in parenthesis are clustered at the case level.
*p<0.1; **p<0.05; ***p<0.01

Table 6: The relationship between initial assignment to PD and being the defendant ranked with the highest predicted sentencing outcome (e.g., prison term, prison)

	All indigents	All multiple	Multiple defendants (PD & CA)		Co-defendants (PD & CA)
	(1)	(2)	(3)	(4)	(5)
Predicted prison term	0.125***	0.124***	0.0365***	0.00580	0.00330
	(0.00129)	(0.00270)	(0.00336)	(0.00381)	(0.00576)
Predicted prison	0.122***	0.117***	0.0338***	0.00588	0.00500
	(0.00124)	(0.00266)	(0.00336)	(0.00382)	(0.00578)
Predicted conviction	0.121***	0.112***	0.0266***	0.00598	0.00347
	(0.00126)	(0.00268)	(0.00336)	(0.00384)	(0.00579)
Predicted trial	0.126***	0.127***	0.0408***	0.0104**	0.00708
	(0.00129)	(0.00270)	(0.00336)	(0.00382)	(0.00577)
obs	651666	182875	84260	84260	35753
PositionFE	No	No	No	Yes	Yes
CaseFE	No	No	Yes	Yes	Yes
DistrictFE	Yes	Yes	No	No	No
YearFE	Yes	Yes	No	No	No

Notes: Each cell in the table reports the coefficient of an indicator whether the defendant was initially assigned a PD or a CA. This is the β coefficient from estimating one of the models (4), (5), or (6). Unlike Table (5), outcome represents an indicator for whether the defendant was ranked as facing the most severe charges according to a certain criterion, which varies by each one of the rows. For example, the cell in the first row and the second column, reports the different between defendants assigned a PD vs. a CA in the probability of being the defendant ranked with the longest expected prison term based on the severity of the charges. Standard errors in parenthesis are clustered at the case level.
*p<0.1; **p<0.05; ***p<0.01

Table 7: San Francisco, the effect of being initially assigned a PD vs. a CA on the case sentencing outcomes

	coefficient of interest Initial PD indicator					
	All indigent		All Multiple		Multiple PD & CA	
	(1)	(2)	(3)	(4)	(5)	(6)
Convicted	-0.070^{***}	-0.016^{***}	-0.029^{***}	-0.037^{***}	-0.039^{***}	-0.039^{***}
	(0.005)	(0.005)	(0.009)	(0.012)	(0.012)	(0.012)
Ave.	0.592	0.592	0.586	0.609	0.609	0.609
Incarcerate	-0.050^{***}	-0.020^{***}	-0.009	-0.012	-0.014	-0.012
	(0.004)	(0.004)	(0.007)	(0.010)	(0.010)	(0.010)
Ave.	0.162	0.162	0.144	0.146	0.146	0.146
Prison	-0.059^{***}	-0.032^{***}	-0.016^{***}	-0.018^{**}	-0.019^{***}	-0.018^{**}
	(0.003)	(0.003)	(0.005)	(0.007)	(0.007)	(0.007)
Ave.	0.068	0.068	0.08	0.079	0.079	0.079
Jail	0.005	0.010^{***}	0.004	0.003	0.003	0.003
	(0.003)	(0.003)	(0.005)	(0.008)	(0.008)	(0.008)
Ave.	0.099	0.099	0.07	0.071	0.071	0.071
asinh(Incarceration term)	-0.357^{***}	-0.180^{***}	-0.099^{***}	-0.108^{**}	-0.109^{**}	-0.104^{**}
	(0.020)	(0.019)	(0.031)	(0.045)	(0.045)	(0.046)
Ave.	0.546	0.546	0.59	0.597	0.597	0.597
asinh(Prison term)	-0.331^{***}	-0.189^{***}	-0.100^{***}	-0.105^{**}	-0.107^{**}	-0.101^{**}
	(0.019)	(0.018)	(0.028)	(0.041)	(0.042)	(0.042)
Ave.	0.359	0.359	0.436	0.435	0.435	0.435
No bail	-0.086^{***}	-0.019^{***}	-0.010	-0.015	-0.018	-0.018
	(0.005)	(0.005)	(0.009)	(0.014)	(0.014)	(0.014)
Ave.	0.338	0.338	0.342	0.359	0.359	0.359
Case FE	No	No	No	Yes	Yes	Yes
Controls	No	Yes	Yes	No	Yes	Yes
Prior PD control	No	No	No	No	No	Yes
Observations	67,613	67,613	9,576	7,164	7,164	7,164

Notes: Each cell in the table contains the coefficient on an indicator whether the defendant was initially assigned a PD or a CA. The standard errors are cluster-robust at the case level, which is the level in which treatment—attorney type—is assigned. Both incarceration and prison terms are measured in months. I approximate the Log(·) function using the asinh(·) function which is a common procedure when the outcome of interest is both skewed and has a mass at zero.

*p<0.1; **p<0.05; ***p<0.01

Table 8: San Francisco: Heterogeneity in the attorney type effect on the primary sentencing outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Any prison			
Initial PD	−0.018**	−0.016	−0.011	−0.020**	−0.013	−0.008	−0.005
	(0.007)	(0.010)	(0.011)	(0.009)	(0.008)	(0.009)	(0.005)
Initial PD × Female		−0.008					
		(0.019)					
Initial PD × Black			−0.012				
			(0.017)				
Initial PD × Hispanic				0.012			
				(0.024)			
Initial PD × Prior incarceration sentence					−0.050		
					(0.034)		
Initial PD × Prior conviction						−0.045*	
						(0.023)	
Initial PD × Any felony in case							−0.015
							(0.010)
				asinh(Prison term)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Initial PD	−0.101**	−0.088*	−0.062	−0.114**	−0.077*	−0.058	−0.027
	(0.042)	(0.053)	(0.059)	(0.051)	(0.044)	(0.049)	(0.030)
Initial PD × Female		−0.056					
		(0.105)					
Initial PD × Black			−0.073				
			(0.092)				
Initial PD × Hispanic				0.061			
				(0.136)			
Initial PD × Prior incarceration sentence					−0.241		
					(0.183)		
Initial PD × Prior conviction						−0.193	
						(0.127)	
Initial PD × Any felony in case							−0.085
							(0.055)
				Convicted			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Initial PD	−0.039***	−0.041***	−0.039**	−0.048***	−0.038***	−0.032**	−0.062*
	(0.012)	(0.015)	(0.019)	(0.015)	(0.013)	(0.015)	(0.032)
Initial PD × Female		0.010					
		(0.037)					
Initial PD × Black			−0.001				
			(0.027)				
Initial PD × Hispanic				0.043			
				(0.041)			
Initial PD × Prior incarceration sentence					−0.015		
					(0.048)		
Initial PD × Prior conviction						−0.030	
						(0.036)	
Initial PD × Any felony in case							0.027
							(0.035)
N	7,164	7,164	7,164	7,164	7,164	7,164	7,164

Notes: Each cell in the table contains the γ coefficient from equation (8) for different dimensions of possible heterogeneity (e.g., gender, criminal history). All coefficient estimates in the table are based only on records from San Francisco. The cluster-robust standard errors are clustered at the case level.
*p<0.1; **p<0.05; ***p<0.01

Table 9: P-values of observed effects in Figure (10)

	asinh(Prison term)	Prison	Convicted
Female	0.468	0.575	0.690
Black	0.275	0.301	0.972
Hispanic	0.555	0.502	0.181
Prior incarceration	0.075	0.051	0.684
Prior conviction	0.032	0.005	0.250
Felony case	0.034	0.033	0.295

Notes: Each cell in the table reports the P-value of the observed effect (red triangular) in Table (10). The P-value is the number of times that a the estimated effect under a random permutation of treatment (black dots) was more extreme than the observed estimated effect.

Table 10: Federal courts, the effect of being initially assigned a PD vs. a CA on the case sentencing outcomes

	All indigent		All multiple		Multiple (PD & CA)		Multiple (PD & CA)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
asinh(Prison term)	-0.278***	-0.0322***	-0.0841***	0.0368***	0.145***	0.115***	-0.0459***	-0.0464***
	(0.00556)	(0.00453)	(0.0106)	(0.00927)	(0.0122)	(0.0119)	(0.0136)	(0.0134)
Prison	-0.00801***	0.00411***	-0.00435*	0.00788***	0.0166***	0.0132***	-0.00811**	-0.00819**
	(0.00100)	(0.000930)	(0.00197)	(0.00189)	(0.00253)	(0.00252)	(0.00288)	(0.00286)
asinh(Probation term)	0.0237***	-0.00522	0.0433***	0.00340	-0.0170*	-0.0113	0.0246**	0.0239*
	(0.00332)	(0.00314)	(0.00640)	(0.00609)	(0.00825)	(0.00821)	(0.00952)	(0.00944)
Some plea	0.0161***	0.00983***	0.0174***	0.0171***	0.0157***	0.0151***	0.00424	0.00432
	(0.000781)	(0.000754)	(0.00171)	(0.00167)	(0.00221)	(0.00221)	(0.00253)	(0.00253)
Trial	-0.0151***	-0.00787***	-0.0136***	-0.0113***	-0.00652***	-0.00782***	-0.00738***	-0.00762***
	(0.000537)	(0.000506)	(0.00123)	(0.00119)	(0.00155)	(0.00155)	(0.00177)	(0.00177)
Num. conviction	-0.0526***	0.0142***	0.0325***	0.0451***	0.0522***	0.0319***	-0.00390	-0.00502
	(0.00182)	(0.00160)	(0.00428)	(0.00392)	(0.00521)	(0.00490)	(0.00586)	(0.00551)
Conviction	0.00282***	0.00284***	0.00642***	0.00825***	0.0113***	0.00959***	-0.00219	-0.00233
	(0.000637)	(0.000627)	(0.00135)	(0.00133)	(0.00179)	(0.00179)	(0.00204)	(0.00203)
Num. dismissed	-0.365***	-0.0229***	-0.136***	-0.0421***	0.0206**	-0.0199***	0.0135	0.0107
	(0.00405)	(0.00195)	(0.00765)	(0.00478)	(0.00659)	(0.00560)	(0.00745)	(0.00632)
Acquittal	-0.00106***	-0.000780***	-0.00191***	-0.00198***	-0.00197***	-0.00184**	-0.000554	-0.000527
	(0.000203)	(0.000195)	(0.000450)	(0.000442)	(0.000584)	(0.000583)	(0.000646)	(0.000646)
Some diversion	0.00120***	0.000537***	0.000224	0.0000213	-0.000347	-0.000273	-0.000147	-0.000106
	(0.000109)	(0.000108)	(0.000179)	(0.000174)	(0.000243)	(0.000240)	(0.000235)	(0.000234)
N	651,666	651,666	182,875	182,875	84,260	84,260	84,260	84,260
Def. Num. FE	No	No	No	No	No	No	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Case FE	No	No	No	No	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	No	No	No	No

Notes: Each cell in the table contains the coefficient on an indicator whether the defendant was initially assigned a PD. The table reports the estimates of the β coefficient from each of the models (4), (5), and (6). Standard errors are clustered-robust at the case level.

*p<0.1; **p<0.05; ***p<0.01

Table 11: San Francisco: Attorney characteristics by attorney type, at the defendant level

	CA	PD	Private
Female	0.253	0.472	0.194
Asian	0.039	0.166	0.036
White	0.573	0.408	0.552
Hispanic	0.031	0.099	0.068
Ave. Rank BA (USnews)	54.661	44.767	51.675
Ave. Rank JD (USnews)	77.015	47.981	65.995
Ave. No rank BA (USnews)	0.169	0.137	0.198
Ave. No rank JD (USnews)	0.895	0.836	0.847
Experience (median)	22.287	6.256	16.776
Num. cases first attorney (median)	39	208	7
Num. cases terminating attorney (median)	56	173	10

Notes: The table shows the characteristics of the initial attorney that represented each defendant. All the calculations in the table where done at the defendant level. The numbers are attorney characteristics averaged across defendants. This is equivalent to the average of attorney characteristics re-weighted by the number of defendants that each individual attorney represented. The "Num. cases first attorney" is the number of cases in which the attorney was the first assigned attorney in a case, and similarly "Num. cases first attorney" is the number of cases in which the attorney was the terminating attorney.

Table 12: San Francisco: The effect of having a PD vs. a CA on the case sentencing outcomes when controlling for attorney characteristics

	Initial PD effect			
	(1)	(2)	(3)	(4)
asinh(Prison term)	-0.118^{***}	-0.119^{***}	-0.101	-0.108
	(0.044)	(0.045)	(0.090)	(0.089)
Prison	-0.021^{***}	-0.021^{***}	-0.015	-0.016
	(0.008)	(0.008)	(0.016)	(0.016)
Convicted	-0.040^{***}	-0.046^{***}	-0.004	-0.012
	(0.014)	(0.014)	(0.025)	(0.025)
Case FE	Yes	Yes	Yes	Yes
Defendant controls	No	Yes	No	Yes
Attorney controls	No	No	Yes	Yes
Observations	6,703	6,703	6,703	6,703

Notes: Each cell in the table contains the coefficient on an indicator whether the defendant was initially assigned a PD or a CA. The standard errors are cluster-robust at the case level. Both incarceration and prison terms are measured in months. I approximate the Log(·) function using the asinh(·) function which is a common procedure when the outcome of interest is both skewed and has a mass at zero. The attorney characteristics include all the covariates in Table (11). The number of observations in this table is smaller than in Table (7), 6703 vs. 7164, since in some of the observations the attorney type was available but the attorney name was either not available or was partially listed.
*p<0.1; **p<0.05; ***p<0.01

Table 13: San Francisco: Changes in attorney characteristics between first and terminating attorneys

	CA	PD	RE
Change attorney	0.207	0.521	0.161
Higher rank JD (US news)	0.010	0.080	0.006
Higher rank BA (US news)	0.079	0.232	0.060
Higher experience	0.163	0.354	0.142
Lower rank JD (US news)	0.011	0.089	0.014
Lower rank BA (US news)	0.091	0.214	0.070
Lower experience	0.100	0.167	0.080

References

Abdulkadiroglu, Atila, Parag A. Pathak, and Christopher R. Walters, "Free to Choose: Can School Choice Reduce Student Achievement?," Working Paper 21839, National Bureau of Economic Research December 2015.

Abrams, David and Albert Yoon, "The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability," *The University of Chicago Law Review*, 2007, *74* (4), 1145–1177.

_ **and Ryan Fackler**, "To Plea or Not to Plea: Evidence from North Carolina," July 2017. .

_ **, Marianne Bertrand, and Sendhil Mullainathan**, "Do Judges Vary in Their Treatment of Race?," *The Journal of Legal Studies*, 2012, *41* (2), 347–383.

Agan, Amanda, Matthew Freedman, and Emily Owens, "Is Your Lawyer a Lemon? Incentives and Selection in the Public Provision of Criminal Defense," June 2017. .

Aizer, Anna and Joseph J. Doyle Jr., "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges *," *The Quarterly Journal of Economics*, 2015, *130* (2), 759–803.

Allison, Junius L., "Relationship Between the Office of Public Defender and the Assigned Counsel System," *Valparaiso University Law Review*, 1976.

Altonji, Joseph, Todd Elder, and Christopher Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Anderson, James and Paul Heaton, "How Much Difference Does the Lawyer Make? The Effect of Defense Counsel on Murder Case Outcomes," *Yale Law Review*, 2012, *122* (1).

Anderson, Michael and Jeremy Magruder, "Split-Sample Strategies for Avoiding False Discoveries," 2017. .

Ater, Itai, Yehonatan Givati, and Oren Rigbi, "The Economics of Rights: Does the Right to Counsel Increase Crime?," *American Economic Journal: Economic Policy*, May 2017, *9* (2), 1–27.

Athey, Susan, Dean Eckles, and Guido W. Imbens, "Exact p-Values for Network Interference," *Journal of the American Statistical Association*, 2018, *113* (521), 230–240.

Bowers, Jake. and Ben B. Hansen, "Attributing Effects to A Cluster Randomized Get-Out-The-Vote Campaign," *Journal of the American Statistical Association*, 2009, *104* (487), 873–885.

Card, David and Stefano DellaVigna, "Nine Facts about Top Journals in Economics," *Journal of Economic Literature*, 2013, *51* (1), 144–161.

_ **, David S. Lee, Zhuan Pei, and Andrea Weber**, "Inference on Causal Effects in a Generalized Regression Kink Design," *Econometrica*, 2015, *83* (6), 2453–2483.

Cheng, Cheng and Wei Long, "Can the Private Sector Provide Better Police Services?," 2017. .

Coase, R. H., "The Nature of the Firm," *Economica*, 1937, *4* (16), 386–405.

Cohen, Alma and Crystal S. Yang, "Judicial Politics and Sentencing Decisions," *accepted, American Economic Journal: Economic Policy*, 2018.

DeFrances, Carol and Marika Litras, "Indigent Defense Services In Large Counties, 1999," NCJ 184932, Bureau of Justice Statistics November 2000.

Dobbie, Will, Jacob Goldin, and Crystal Yang, "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," Working Paper 22511, National Bureau of Economic Research August 2016.

Farole, Donald and Lynn Langton, "County-based and Local Public Defender Offices, 2007," NCJ 231175, Bureau of Justice Statistics November 2010.

_ and _ , "State Public Defender Offices, 2007," NCJ 228229, Bureau of Justice Statistics November 2010.

Field, Erica, "Educational Debt Burden and Career Choice: Evidence from a Financial Aid Experiment at NYU Law School," *American Economic Journal: Applied Economics*, 2009, *1* (1), 1–21.

Fischman, Joshua B., "Estimating Preferences of Circuit Judges: A Model of Consensus Voting," *The Journal of Law and Economics*, 2011, *54* (4), 781–809.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo, "Decomposition Methods in Economics," in O. Ashenfelter and D. Card, eds., *O. Ashenfelter and D. Card, eds.*, 1 ed., Vol. 4A, Elsevier, 2011, chapter 01, pp. 1–102.

Gelbach, Jonah B., "When Do Covariates Matter? And Which Ones, and How Much?," *Journal of Labor Economics*, 2016, *34* (2), 509–543.

Gelber, Alexander, Adam Isen, , and Jae Song, "The Effect of Pension Income on Elderly Earnings: Evidence from Social Security and Full Population," 2016.

Hart, Oliver, Andrei Shleifer, and Robert W. Vishny, "The Proper Scope of Government: Theory and an Application to Prisons," *The Quarterly Journal of Economics*, 1997, *112* (4), 1127–1161.

Haugh, John J., "The Federal Criminal Justice Act of 1964: Catalyst in the Continuing Formulation of the Rights of the Criminal Defendant," *Notre Dame Law Review*, 1966, *41* (6).

Iyengar, Radha, "An Analysis of the Performance of Federal Indigent Defense Counsel," Working Paper 13187, National Bureau of Economic Research June 2007.

Joy, Peter and Kevin Mcmunigal, "Does the Lawyer Make a Difference? Public Defender v. Appointed Counsel," 05 2012.

Judicial-Conference, "Report of the Judicial Conference of the United States," Technical Report, Administrative Office of the United States Courts 1952.

Keele, Luke and Luke Miratrix, "Randomization Inference for Outcomes with Clumping at Zero," *The American Statistician*, 2017, *0* (ja), 0–0.

Kuziemko, Ilyana, "How Should Inmates Be Released From Prison? An Assessment of Parole Versus Fixed Sentence Regimes," *Quarterly Journal of Economics*, 2013, *128* (1), 371–424.

Leacy, Finbarr P. and Elizabeth A. Stuart, "On the joint use of propensity and prognostic scores in estimation of the average treatment effect on the treated: a simulation study," *Statistics in Medicine*, 2014, *33* (20), 3488–3508.

Lowenthal, Gary T., "Joint Representation in Criminal Cases: A Critical Appraisal," *Virginia Law Review*, 1978, *64* (7), 939–989.

Moore, Stephen C., "Conflicts of Interest in Public Defender Offices," *Journal of the Legal Profession*, 1984.

Mueller-Smith, Michael, "The Criminal and Labor market Impacts of Incarceration," Working Paper 2015.

Mukherjee, Anita, "Impacts of Private Prison Contracting on Inmate Time Served and Recidivism," 2017. .

Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 1973, *14* (3), 693–709.

Paetzold, Jörg and Hannes Winner, "Bunching in the presence of deduction possibilities: Earnings and deduction responses to a large kink," Technical Report 2016.

Prado, Edward, "Process and Progress: Reviewing the Criminal Justice Act," *Law and Contemporary Problems*, 1995, *58* (1).

_ , Robert Altman, Vincent Aprile, Judy Clark, Michael Davis, Edward Dennis, Robinson Everett, Thomas Hillier, George Revercomb, and Charles Ogletrel, "Report of the Committee to Review the Criminal Justice Act," Technical Report, United States Judicial Conference 1993.

Raphael, Steven, *Controlling Crime: Strategies and Tradeoffs*, University of Chicago Press, 2011.

Rehavi, Marit and Sonja Starr, "Racial Disparity in Federal Criminal Sentences," *Journal of Political Economy*, 2014, *122* (6), 1320–1354.

Roach, Michael A., "Indigent Defense Counsel, Attorney Quality, and Defendant Outcomes," *American Law and Economics Review*, 2014.

Schwall, Benjamin, "High-Powered Attorney Incentives: A Look at the New Indigent Defense System in South Carolina," Technical Report, Social Science Research Network 2015.

Silveira, Bernardo S., "Bargaining With Asymmetric Information: An Empirical Study of Plea Negotiations," *Econometrica*, 2017, *85* (2), 419–452.

Yang, Crystal, "Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing," *The Journal of Legal Studies*, 2015, *44* (1), 75–111.

Online supplementary appendix

A Supplementary figures and tables

Figure A.1: Federal courts, the distribution of defendants across attorney types and over time, by filing year

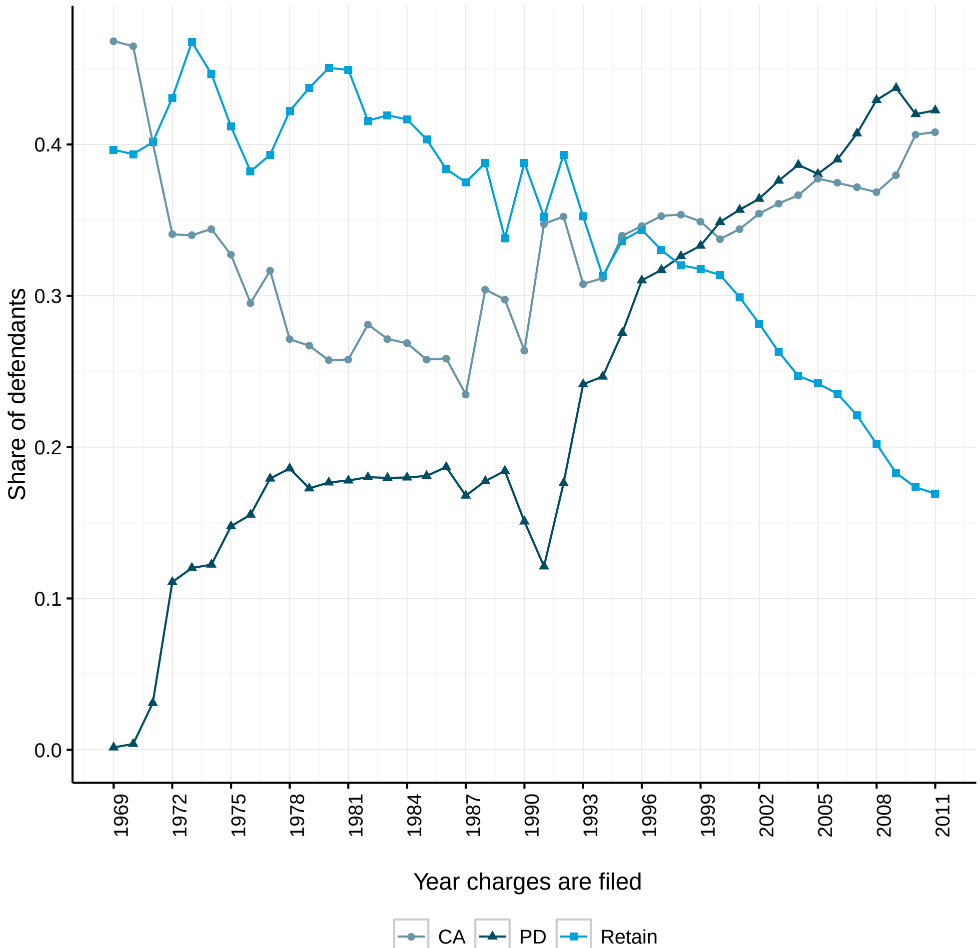


Figure A.2: federal courts, validating the conflict of interest hypothesis (1970 – 1995)

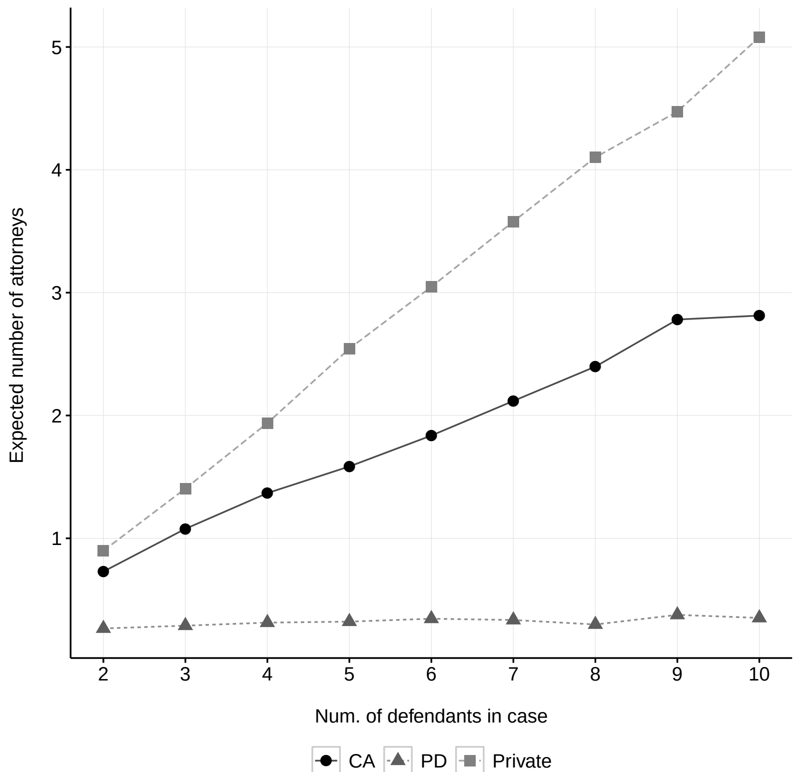
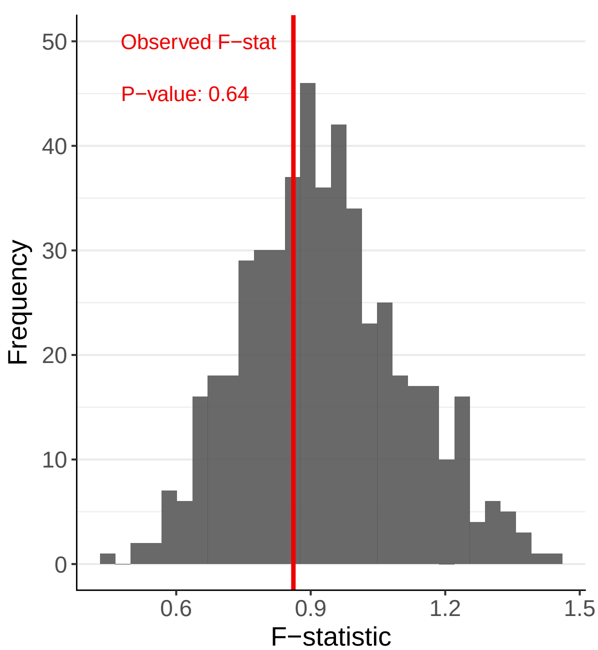
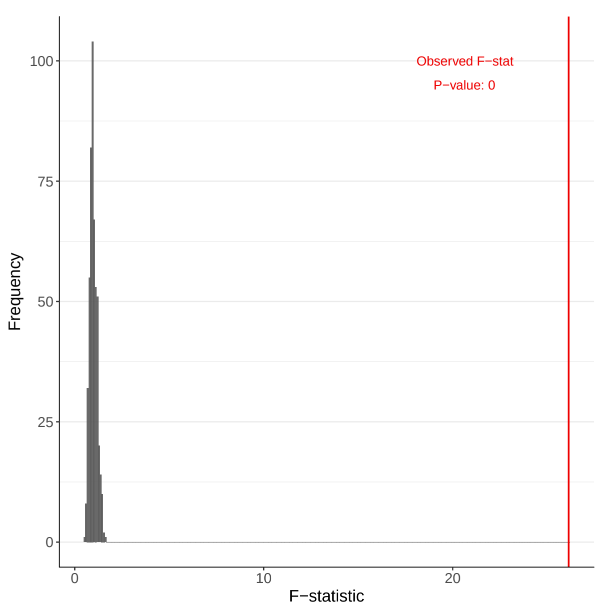


Figure A.3: Monte-Carlo permutations of attorney type assignment within a case: F-statistic using Offense codes



(a) Within multiple defendant case



(b) Across single defendant cases

Notes: See Figure (A.3)

Table A.1: Descriptive statistics on indigent defendants in federal courts ($1970 - 1995$)

	All	Single def.	Multiple def.	Multiple def. (PD & CA)
	(1)	(2)	(3)	(4)
Prison term (months)	27.600	21.800	37.260	35.340
Prison	0.520	0.500	0.570	0.590
Conviction	0.830	0.850	0.800	0.820
Plea	0.730	0.770	0.660	0.700
Trial	0.130	0.100	0.170	0.150
Acquittal	0.030	0.020	0.030	0.030
Predicted prison term	28.920	24.360	36.500	36.690
Predicted prison	0.530	0.500	0.580	0.600
Predicted probation term	67.650	71.240	61.670	50.380
Predicted plea	0.710	0.720	0.700	0.710
Predicted trial	0.140	0.130	0.160	0.160
Predicted conviction	0.820	0.820	0.830	0.830
Predicted dismissed	0.170	0.170	0.170	0.160
felony	0.470	0.420	0.570	0.690
Number of defendants	2.380	1	4.690	5.210
CA in case	0.720	0.600	0.900	1
PD in case	0.400	0.400	0.400	1
Observations	494,822	309,083	185,739	50,036

Notes: The table presents descriptive statistics for all criminal defendants in cases terminated in federal district courts from 1970 to 1995. The four columns refer to different sub-samples of the data. The column (1) makes no restrictions and includes all defendants. Column (2) restrict attention to individuals in single defendant cases, without any co-defendants. Column (3) Restrict the sample to individuals in cases that includes more than one defendant. Column (4) restrict the sample in column (3) to multiple defendant cases in which at least one defendant is represented by a PD and another by a CA. In each of these cases there are both defendants who are represented by a PD and a CA, which allows to conduct a within case comparison of attorney types. All the predicted variables are summary measures for the charges that have been filed against the defendant based on a Oaxaca decomposition. Appendix (D) describes how the predicted variables are constructed.

Table A.2: Variation in defendant characteristics within a multiple defendant case

	Multiple	Co-defendants
Obs.	2.143	2.000
Black & Non-Black	0.222	0.214
Hispanic & Non-Hispanic	0.333	0.314
White & Non-White	0.245	0.236
Black & White	0.223	0.217
Black & Hispanic	0.073	0.067
White & Hispanic	0.081	0.073
Felony & Non-Felony	0.014	0.012
Prior arrest & No prior arrest	0.338	0.322
Prior conviction & No prior conv.	0.156	0.158
Prior incarceration & No prior incar.	0.287	0.280

B Detecting random assignment of defendants across attorney types using a data driven procedure in federal district courts

Iyengar (2007) proposed a data-driven procedure to detect location-year pairs in which the assignment of defendants to attorney types—PD or CA—was done at random. A location-year pair is classified as using a random assignment mechanism if the covariates are not predictive of the treatment allocation. More specifically, consider the following model:

$$PD_i = X_i'\alpha + \gamma_{d(i)} + \eta_{t(i)} + \epsilon_i \tag{9}$$

and under random assignment of treatment the covariates should *not* be predictive of assignment to a PD:

$$H_0: \ \alpha = 0 \tag{10}$$

The procedure proposed by Iyengar (2007) is to conduct the hypothesis test in equation (10) in each district-year pair and if the F-statistic is below a certain threshold and is not statistically significant then to classify that district-year pair as using random assignment of defendants across attorney types. The above is a slight variation on the algorithm used by Iyengar (2007), since I am using ordinary-least-square instead of a Probit regression.

I find that no district passes a threshold of $F - stat < 0.5$ for two consecutive years; and 29 districts pass a threshold of $F - stat < 1$ for two consecutive years but only six pass this threshold for three consecutive years and no district passes the threshold for four consecutive years. It is not likely that districts frequently change the assignment procedure and therefore it is not clear whether the algorithm leads to false classification of districts as using random assignment or perhaps there is an over-rejection problem.

Next covariate balance is assessed in the sample of districts that passes the above procedure and are classified as using a random allocation procedure. The following econometric model is used to asses the covariate balance:

$$X_i = \beta \cdot PD_i + \gamma_{d(i)} + \eta_{t(i)} + e_i \tag{11}$$

where γ_d and η_t are district and filing year fixed effects. This model is the same as model (4); and the β coefficient can be interpreted as the difference in means in characteristic X_i between defendants who have been assigned a PD relative to a CA.

Since we include district and filing year fixed effects the covariates should not be predictive of the attorney type—PD or CA—assignment. Table (B.1) reports the results for several different thresholds of the F-statistic. In column (1), all the balance tests look good; however, no district passes this threshold for two consecutive years. In column (2) there are significant imbalances and the differences increase as the $F - stat$ threshold is increased to 1.5 (column 3).

Table B.1: Tests for whether charge severity measures predict initial PD assignment among districts with imputed "random" allocation of PDs

	Initial assignment public defender		
	$F - stat < 0.5$	$F - stat < 1$	$F - stat < 1.5$
	(1)	(2)	(3)
Predicted prison term	−0.001	−0.0002	−0.0002*
	(0.0004)	(0.0001)	(0.0001)
Predicted prison	0.004	−0.190	−0.452***
	(0.281)	(0.121)	(0.082)
Predicted probation term	−0.002	−0.003	−0.009***
	(0.008)	(0.003)	(0.002)
Predicted plea	−0.046	1.252**	1.597***
	(1.043)	(0.599)	(0.433)
Predicted trial	0.306	1.518***	1.779***
	(0.814)	(0.486)	(0.362)
Predicted conviction	−0.116	−0.638	−0.464
	(1.047)	(0.590)	(0.419)
Predicted num. convictions	0.076	0.085***	0.080***
	(0.076)	(0.032)	(0.016)
Predicted num. dismissed	0.021	−0.048**	−0.103***
	(0.050)	(0.022)	(0.014)
Observations	2,197	21,500	56,721

Notes: The table includes also dummy variables for the number of felony charges the defendant is charged and have been removed from the table due to space limitations. Robust standard errors in parenthesis. The data includes only single defendant cases with at least one felony level charge.
*p<0.1; **p<0.05; ***p<0.01

C The history of the right to appointed-counsel in the U.S.

The Sixth Amendment in the Bill of Rights states that "[i]n all criminal prosecutions, the accused shall enjoy the right... to have the Assistance of Counsel for his defense." However, the Constitution leaves open the question of what should happen when a defendant cannot afford to hire an attorney. In 1932, the Supreme Court ruled (*Powell v. Alabama*) that defendants charged with capital cases in state and federal courts who cannot afford an attorney have a constitutional right to have one appointed by the court. In 1938, the Supreme Court extended *Powell* and ruled that federal defendants in *all* felony criminal cases have a right to an appointed counsel (*Johnson v. Zerbst*). However in 1942, the Supreme Court decided in *Betts v. Brady* that *Johnson v. Zerbst* did not extend to defendants charged with non-capital cases at the state level. Although federal criminal defendants had the right to an appointed counsel since 1938, the question of who will compensate the appointed counsel remained open, and the provision of professional legal counsel to federal low-income defendants was limited (Judicial-Conference, 1952).

The Supreme Court established the right of indigent defendants to a court-appointed counsel in the 1960s. In 1963, the landmark ruling of the Supreme Court in *Gideon v. Wainwright* extended the limited scope of the *Powell* and *Johnson* decisions when it overturned *Betts v. Brady* by requiring states to provide a legal counsel to defendants facing any felony charges. In 1972, these rights were further extended to all criminal prosecutions that carry a sentence of imprisonment in

Argersinger v. Hamlin. In *Scott v. Illinois*, the Supreme Court interpreted *Argersinger v. Hamlin* as referring only to a sentence of *actual* imprisonment. It determined that the criterion for whether a defendant is entitled to a court-appointed counsel is whether he was sentenced to an actual period of incarceration. In 2002, *Shelton vs. Alabama* extended the right to a court-appointed counsel also to defendants who are sentenced to a suspended sentence of incarceration (e.g., probation).[21]

In 1964, the Criminal Justice Act assured federal defendants professional legal counsel by establishing a federal indigent defense system financed by the court. The CJA secured compensation for court-appointed attorneys and provided indigent defendants with funds for investigative and expert services to guarantee an adequate defense. In 1970, the law was amended to allow courts to establish federal public defender organizations (Prado, 1995; Haugh, 1966).[22]

[21]The Supreme Court mentioned two conditions under which a defendant facing a suspended sentence of imprisonment will not be eligible to a court-appointed counsel. The first is if the state offers an opportunity to re-litigate the guilt or innocent of the defendant in any future revocation proceedings. The second scenario is that the probation term cannot be revoked and replaced by actual imprisonment.

[22]The Criminal Justice Act allows several districts to share the services of a single Public Defender Office, as long as their cumulative number of cases is at least 200.

D Covariate indices for charge severity measures

To quantify the gaps in the severity of the filed charges between defendants that are assigned a PD and those assigned a CA, I consider a simple summary measure of the selection based on a Oaxaca decomposition. A trial outcome (e.g., incarceration length) Y_{ig} can be modelled by projecting it on a set of pre-trial charge characteristics:

$$Y_{ig} = X_{ig}\beta_g + \nu_g, \quad \text{where} \quad g = \text{PD, CA} \tag{12}$$

The coefficient vector β_g has a causal interpretation under certain conditions (Fortin et al., 2011), and the fitted values $X_g\hat{\beta}_g$ are independent of $\hat{\nu}_g$ by construction. The average difference in the trial outcome, $\bar{Y}_{\text{PD}} - \bar{Y}_{\text{CA}}$, between attorney types can be written as (Oaxaca, 1973),

$$\bar{Y}_{\text{PD}} - \bar{Y}_{\text{CA}} = \hat{\beta}_{\text{CA}}\left(\bar{X}_{\text{PD}} - \bar{X}_{\text{CA}}\right) + \left(\hat{\beta}_{\text{PD}} - \hat{\beta}_{\text{CA}}\right)\bar{X}_{\text{PD}} \tag{13}$$

The first element in (13), $\hat{\beta}_{\text{CA}}\left(\bar{X}_{\text{PD}} - \bar{X}_{\text{CA}}\right)$, is the average difference in charge characteristics re-weighted by the effect of each characteristic on the trial outcome among defendants who are represented by a CA. This term represents selection on observables and will be zero in a standard balance test when:

$$\bar{X}_{\text{PD}} = \bar{X}_{\text{CA}} \tag{14}$$

One can summarize the imbalance in initial charge characteristics by estimating the difference in covariate indices $X_i'\hat{\beta}_{\text{CA}}$ that reduces the dimension of the covariate vector X_i to a single dimensional index. The idea of summarizing imbalance by the covariates' relationship to the outcome surface has been proposed in the past by several different procedures (Bowers and Hansen, 2009; Card et al., 2015; Paetzold and Winner, 2016; Leacy and Stuart, 2014).

In San Francisco, I use the covariate index, $X_i'\hat{\beta}_{\text{PD}}$, which is based on estimating β using only defendants that have been assigned a PD. More specifically, I regress each case outcome on a vector of charge, case and defendant characteristics such as demographic characteristics, criminal history, charge severity (e.g., felony, misdemeanor). The main covariates are listed in Table (4) and Figure (4). In addition, I use SC and BCS codes which are 2-digit and 3-digit classifications of offenses to broader categories.[23]

In federal courts, I follow the procedure that as was described above and use the covariate index, $X_i'\hat{\beta}_{\text{CA}}$.[24] More specifically, I regress each case outcome on a vector of charges characteristics such

[23] The classification is done by the California Department of Justice, `https://oag.ca.gov/law/code-tables`.

[24] In federal courts throughout the sample there exists a large fraction of defendants that have been assigned CAs; however, in San Francisco the fraction of the defendants that have been assigned CA is much smaller then those who have been assigned PD. In federal courts, in some districts there is no PD office for part of the period that is why in federal courts I use $\hat{\beta}_{\text{CA}}$ and in San Francisco $\hat{\beta}_{\text{PD}}$. If I use the same index in both the results are the same.

as indicators for the charges the defendants is facing based on a four-digit offense code of the Federal Administrative Office of the Courts.[25] The four-digit codes have many values and hence displaying balance tests for indicators of each of the offense codes is not feasible. This is one of the motivations to use the above dimension reduction. I also include indicators for the number of charges at each severity level (e.g., misdemeanor, felony). In federal district courts, I do not observed criminal history and demographic information about the defendant, unlike San Francisco in which this information is available.

This choice has no implication on the results reported in the paper.

[25] For more details on the four digit codes see https://www.fjc.gov/sites/default/files/idb/codebooks/Criminal%20Code%20Book%201996%20Forward_0.pdf

E Monte Carlo simulations of exact inference vs. cluster-robust standard errors

Figure (9) suggest that conducting inference over the null that the attorney type has no effect using Fisherian inference, also known as exact/permutation inference, can have higher power to reject the null of no effect when it is false. To better understand the power of the two different methods of conducting inference, I conduct a simple Monte-Carlo simulation that examines the performance of the two procedures with respect to power in a data generating process that resembles to the observed data.

Consider the following data generating process of a constant treatment effect. I use the multiple defendants analysis sample and define the observed value of $asinh(\text{Prison term})_i$ as $Y_i(0)$. The potential outcome under the treatment regime is:

$$Y_i(1) = Y_i(0) + \tau \cdot \text{PD}_i \tag{15}$$

Next I randomly assign defendants to treatment regimes (PD_i) using randomization within a multiple defendant case. I use 16 different values of τ and for each one conduct 1,000 random assignments of defendants to treatment and for each such assignment calculate the observed value of Y_i:

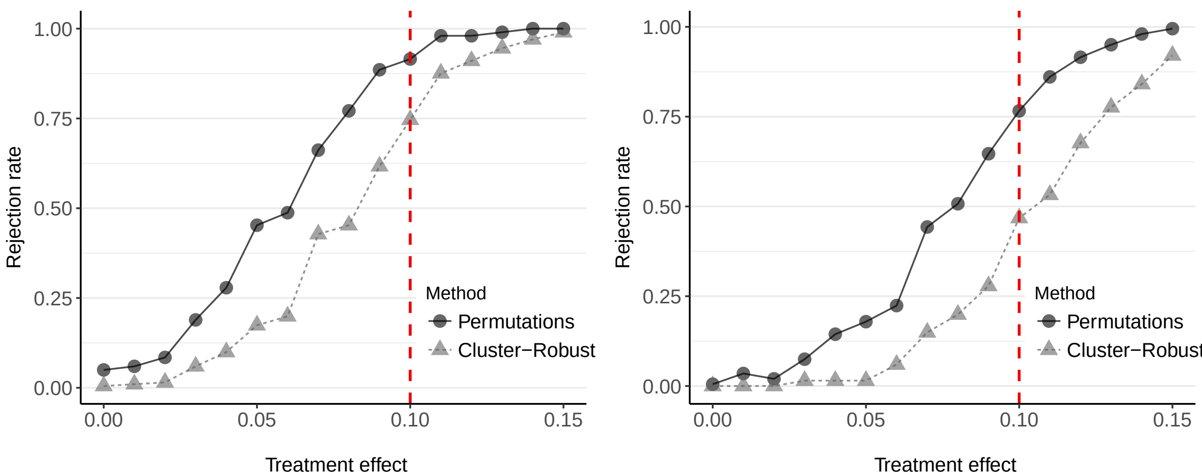
$$Y_i = Y_i(0) \cdot (1 - \text{PD}_i) + Y_i(1) \cdot \text{PD}_i \tag{16}$$

I then conduct inference over the null that τ is different than zero using both permutation inference and the cluster-robust standard errors. The estimator for τ is:

$$Y_i = \tau \text{PD}_i + \alpha_{j(i)} + e_i \tag{17}$$

and when using permutation inference the test statistic is the t-statistic of $\hat{\tau}$ divided by the cluster-robust standard error both are calculated in each random permutation of treatment assignment. Figure (E.1) reports the simulation results in terms of the rejection rate of the two procedures for each one of the values of τ. For higher values of τ both procedures reject the null in a higher rate, however, the permutation inference procedure seems to have a higher rejection rate for every τ and stochastically dominates the cluster-robust inference in terms of power. This Monte-Carlo simulation is specific to this data application and should not be used to make general claims on the efficiency of permutation inference relative to regular inference based on cluster-robust standard errors.

Figure E.1: **The rejection rate using permutation inference compare to the standard cluster-robust standard errors**



(a) Rejection rate at 5% confidence level (b) Rejection rate at 1% confidence level

Notes: The clustering, in the cluster-robust standard errors, is performed at the case level. This is the same level in which the randomization is conducted in the permutation inference calculations.