The Effect of Incarceration on Re-Offending: Evidence from a Natural Experiment in Pennsylvania[1]

Daniel S. Nagin, Carnegie Mellon University[2]

G. Matthew Snodgrass, Carnegie Mellon University

January 16, 2012

Abstract

        This paper uses a sample of U.S. offenders from Pennsylvania to estimate the effect of incarceration on post-release criminality. To do so, we capitalize on a feature of the criminal justice system in Pennsylvania—the county-level randomization of cases to judges. We begin by identifying five counties in which there is substantial variation across judges in the uses of incarceration, but no evidence indicating that the randomization process had failed. After demonstrating covariate balance that is consistent with randomization, we estimate a series of confidence intervals for the effect of incarceration on one year, two year, five year, and ten year rearrest rates. On the whole, there is little evidence in our data that incarceration impacts rearrest.

---

[2] Authors listed alphabetically

INTRODUCTION

The principle aim of this research is to estimate the effect on reoffending of the *experience* of a custodial sanction involving imprisonment as compared to a noncustodial sanction. In considering the crime prevention effects of imprisonment, it is important to differentiate between the potential preventive effect of the *threat* of imprisonment and the potential preventive effect of the *experience* of imprisonment. In criminology, the former is referred to as general deterrence. The *experience* of imprisonment may affect reoffending by either altering an individual's opportunity to reoffend or by altering an individual's behavior when released. The first of these is referred to in the criminological literature as incapacitation, the second as specific deterrence. The logic of incapacitation and the direction of its effect are clear. With respect to the specific deterrent effect, there are many sound theoretical arguments for predicting that the experience of imprisonment will be criminogenic, not preventive.

A recent review of the empirical literature on the effect of the experience of imprisonment compared to noncustodial sanctions concludes that on balance the evidence points to a null or criminogenic effect rather than a preventive effect (Nagin, Cullen, & Jonson, 2009). However, this review also concludes the evidentiary base for this conclusion is weak. Nagin et al. (2009) identify a long list of deficiencies in the research on the effect of imprisonment on reoffending. Included among these deficiencies are insufficient controls for age, prior record of offending, and offense severity, all of which may bias the imprisonment effect estimate. Perhaps the most important limitation of existing research is the selection problem—even in studies with extensive controls for measured differences between individuals who do and do not receive custodial sanctions, unmeasured differences that are systematically related to recidivism probability may be biasing the estimate of the effect of imprisonment on reoffending. To overcome the selection problem, we

capitalize on the random assignment of cases to judges in the criminal courts of Pennsylvania. Because of random assignment, there will be no systematic difference in case characteristics across judges. Cross-judge variation in punitiveness, which we demonstrate exists in Pennsylvania, is used as the basis for inferring the effect of incarceration compared to a noncustodial sanction on reoffending.

BACKGROUND

As stated above, the rationale for incapacitation is clear cut. By incarcerating an individual, his or her opportunity to offend against the community is all but removed[3]. Through incapacitation, society averts, on average, λ crimes per year that the individual would have committed had s/he been free, where λ is the individual's expected rate of offending. While as a logical matter we know λ is non-negative[4], estimates of the size of the effect vary considerably. For example, Sweeten and Apel (2007) use matching find that incapacitation averts approximately 6 crimes per year in their sample of 18-19 year old offenders. In contrast, Owens (2009), exploiting a change in policy, finds that incapacitation averts just under 3 crimes per year.

The logic of specific deterrence is grounded in the idea that the experience of imprisonment will deter reoffending, perhaps because the experience is more adverse than anticipated. Moreover, because the criminal law commonly prescribes more severe penalties for recidivists[5], the structure of the law itself may also cause previously convicted individuals to revise upward their estimates of the likelihood and/or severity of punishment for future law breaking. The experience of punishment may

---

[3] The offender could still engage in crimes that do not require physical proximity to the victim (e.g., fraud, extortion, identity theft). The offender may also still engage in criminal behavior directed toward fellow inmates, correctional employees, or the correctional institution.

[4] Ignoring the rare offense committed while imprisoned, this is true at the individual level. However, the number of crimes averted at the *societal* level needn't be non-negative. If the incapacitated individual is replaced by an individual with a larger λ, then overall levels of crime may rise.

[5] For example, sentencing guidelines routinely dictate longer prison sentences for individuals with prior convictions. Similarly, it may also be the case that prosecutors are more likely to prosecute individuals with criminal histories.

also affect the likelihood of future crime by decreasing the attractiveness of crime itself or by expanding alternatives to crime. While imprisoned the individual may benefit from educational or vocational training that increases post-release non-criminal income earning opportunities (Layton MacKenzie, 2002). Other types of rehabilitation are designed to increase the capacity for self-restraint when faced with situations, like a confrontation, that might provoke a criminal act such as violence (Cullen, 2002).

On the other hand, there are many reasons for theorizing that the experience of punishment might increase an individual's future proclivity for crime. While some individuals might conclude that imprisonment is not an experience to be repeated, others might conclude that the experience was not as adverse as anticipated and as a result be more, not less, crime prone. Prisons might be 'schools for crime' where inmates learn new crime skills even as their non-crime human capital depreciates. Associating with other more experienced inmates could lead new inmates to adopt the older inmates' deviant value systems or enable them to learn 'the tricks of the trade' (Adams, 1996; Hawkins, 1976; Steffensmeier & Ulmer, 2005). Being punished could also elevate the offender's feelings of resentment against society (Sherman, 1992) or strengthen the offender's deviant identity (Matsueda, 1992).

The experience of imprisonment may also increase future criminality by stigmatizing the individual socially and economically. There is much evidence showing that an important part of the deterrent effect of legal sanctions stems from the expected societal reactions set off by the imposition of legal sanctions (Williams and Hawkins, 1986; Nagin and Pogarsky, 2003; Nagin and Paternoster, 1994). Prior research has found that individuals who have higher stakes in conformity are more reluctant to offend when they risk being publicly exposed (Klepper and Nagin, 1989). While the fear of arrest and stigmatization may deter potential offenders from breaking the law,

those that have suffered legal sanctions may find that conventional developmental routes are blocked. In their work on the 500 Boston-delinquents initially studied by Glueck and Glueck (1950), Sampson and Laub (1997) have called attention to the role of legal sanctions in what they call the process of cumulative disadvantage. Official labeling through legal sanctions may cause the offender to become marginalized from conventional opportunities and non-criminal social networks, which in turn increases the likelihood of their subsequent offending (Bernburg and Krohn, 2003). Sampson and Laub (1997) propose that legal sanctions may amplify a 'snowball' effect that increasingly 'mortgages' the offender's future by reducing conventional opportunities. Several empirical studies support the theory that legal sanctions downgrade conventional attainment (Freeman, 1996; Nagin and Waldfogel, 1995, 1998; Sampson and Laub, 1993; Waldfogel, 1994; Western, 2002; Western, Kling & Weiman, 2001) and increase future offending (Bernburg and Krohn, 2003; Hagan and Palloni, 1990).

Although space does not permit an extended discussion of the evidence on the effect of imprisonment on reoffending, there are a few observations that deserve mention. First, in terms of numbers, the great majority of studies based on non-experimental data point to a criminogenic effect of custodial sanctions compared to non-custodial sanctions (Nagin et al., 2009). As already indicated, much of this research is vulnerable to the criticism that persons sent to prison are more crime prone in unmeasured ways and as a result, the seeming criminogenic effect of imprisonment is entirely or at least substantially attributable to "selection" bias.

Second, there have been a small number of experimental or quasi-experimental studies comparing custodial v. non-custodial sanctions. Nagin et al. conclude that, taken as a whole, the experimental studies also point towards a criminogenic effect of custodial sanctions. The evidence for this conclusion, however, is weak because it is based on only a small number of studies and

many of the point estimates are not statistically significant. Further, several features of the samples used in these studies also limit their usefulness for understanding the effects of imprisonment on reoffending in the contemporary context of imprisonment in America. Of the five experimental examinations, two involve juveniles and all but one (Killias, Aebi, & Ribeaud, 2000) utilize data that is more than 20 years old. Among the four studies involving adults, only Bergman (1976) is based on a population that might be characterized as serious adult offenders[6].

Three other studies are also notable for our proposed research – Berube and Green (2007), Green and Winik (2010) and Loeffler (2011). Similar to this work, each of these studies use the random assignment of cases to judges to overcome the selection problem. None of these studies found evidence that the experience of imprisonment affected reoffending. Our proposed research moves beyond these valuable efforts in several ways. First, as elaborated upon in the section 3, random case assignment guarantees that any difference across judges in the recidivism rates of their case loads is attributable to a "judge" effect. Thus, an important first step in the analysis is establishing that "judge" treatment effects are present and large, something that was only done in the Loeffler analyses. Second, we extensively check balance across judges in observed covariates. Third, instead of relying on the output from an instrumental variable regression, our analysis takes a different approach developed by Rosenbaum and colleagues (Imbens and Rosenbaum, 2005; Rosenbaum, 1996, 1999, 2002a, 2002b).

While not necessarily better than classic econometric techniques for estimating the impact of treatment using instrumental variables, our approach adds value in two important ways. First, our approach develops an individual-level model of the response to incarceration. Rather than relying

---

[6] The subjects of the Killias, Aebi, and Ribeaud (2000) and the Van der Werff (1979) studies both involved populations of individuals who, if imprisoned, would have received sentences of 14 days or less. Thus, their offenses were unlikely to be very serious.

on stochastic disturbances in a regression framework, this approach clearly develops a counterfactual argument and then relates our inference back to the counterfactual model. Second, and more importantly, our approach generates statistically valid confidence intervals even when the instrument is uninformative or weak. The problem of weak instruments is well known and well documented (Bound, Jaeger, and Baker, 1995; Nelson and Startz, 1990; Maddalla and Jeong, 1992). The classic Two-Stage Least Squares (TSLS) approach to the estimation of treatment effects using instrumental variables relies on asymptotic properties. However, the finite sample and asymptotic properties on which TSLS rely are highly questionable when the instrument is weak or uninformative (Nelson and Startz, 1990; Maddalla and Jeong, 1992). Rather than obfuscating the limitations associated with a weak or uninformative instruments, our approach continues to yield valid confidence intervals even when the instrument is weak. Specifically, rather than providing a point estimate and associate standard error that is driven by incorrect asymptotics, our approach yields confidence intervals that grow in length as the information contained in the data degrades. Put differently, the approach taken in this paper will inform how, rather than assume that, the data is informative concerning the treatment effect.

DATA

To estimate the impact of incarceration on subsequent criminality, this study uses a sample of 6,515 offenders *convicted* of a criminal offense in the Court of Common Pleas in the state of Pennsylvania during 1999 who had their sentencing information forwarded to the Pennsylvania Sentencing Commission. As discussed in greater detail below, for each of these 6,515 offenders we observe basic demographic characteristics, extent and severity of offending history, seriousness of current offense, the type of punitive sanction (e.g., release to the community, sentence to county jail,

sentence to state prison), and the duration of the incarceration administered by the judge. To measure future offending, we observe any arrest that occurred in the state of Pennsylvania.

We use six Pennsylvania counties to estimate the impact of incarceration on rearrest. To be clear, in Pennsylvania, randomization occurs at the level of the county—the geographic unit to which Common Pleas judges are elected. Among the duties of Common Pleas judges is the adjudication and sentencing of criminal cases. Pennsylvania is composed of sixty-seven counties. The number of judge elected in a county depends upon its population. This analysis began by identifying Pennsylvania counties that satisfied three conditions. First, pre-sentence covariates were examined to identify counties in which the 1999 randomization process achieve balance on observable covariates[7]. Second, sentencing outcomes were examined across time to identify counties in which the judiciary demonstrated stable sentencing practices. Third, conditional on satisfying the previous two requirements, counties in which there was statistically significant variation across judge in the use of confinement, whether in the form of jail or prison, were selected[8]. This process lead to the identification of six counties that satisfied that satisfied the selection criteria – Centre, Crawford, Cumberland, Dauphin, Erie, and Mercer Counties.[9] Figure 1 displays a county map of Pennsylvania with these six counties highlighted in red.

****

Figure 1. Map of Pennsylvania with Selected Counties

****

---

[7] Balance on observables is shown in section 5.1
[8] Variation in the use of confinement is shown in section 5.2. We also selected one county in which there was no variation in the use of confinement as a control.
[9] Pennsylvania's two largest counties, Allegheny and Philadelphia, assign a subset of judges to hear specific types of cases usually involving less serious charges. Our data do not identify judges who were so assigned. As a consequence, it was not possible to test for balance and differences in judge punitiveness in these two counties. For this reason they were not candidate counties for inclusion in the analysis.

As mentioned above, in these six counties 6,515 offenders were convicted of a criminal offense in the Court of Common Plea during 1999 and had their sentencing information forwarded to the Pennsylvania Sentencing Commission[10]. The randomization of cases occurs when the case is docketed[11], which is prior to conviction. Thus, we do not observe cases that were randomly assigned to judges but did not result in conviction. From the 6,515 offenders convicted in these six counties, we set aside data from 110 (1.7%) offenders who were sentenced by a judge who did not sentence at least 100 offenders in 1999. This restriction was made to ensure that the observed sentencing outcomes reflect the judge's true underlying tendency to mete out incarceration. We use the information contained in these 6,405 offenders' cases to verify that the 1999 randomization achieved the desired level of balance in the pre-sentence and case disposition covariates.

The pre-sentence covariates and case disposition measures used in this analysis were supplied by the Pennsylvania Sentencing Commission (PASC). PASC data allows the tracking of cases from sentencing through release or entry into the correctional system, depending on sentencing outcome. At the case level, PASC data documents the county of adjudication, the judge of record, and the number of charges in each case. PASC data also records basic demographic information including age, sex, and race. Additionally, the data permits observation of each offender's prior record score[12], and the number of prior adjudications and convictions for 25 separate offense categories (e.g., number of prior burglary offenses, number of prior rape offenses, number of prior

---

[10] "Not all sentences are reported to the Commission. 1) Philadelphia Municipal Court sentences are not reported to the Commission. These may include DUI (driving under the influence) offenses as well as other misdemeanor offenses. 2) Offenses sentenced by district magistrates are not reported to the Commission. These typically include DUI offenses or other misdemeanor offenses. 3) Murder 1 and Murder 2 offenses, which are subject to life or death mandatory sentences, do not fall under the sentencing guidelines and are not required to be reported to the Commission. The Commission encourages reporting of the Murder 1 and Murder 2 offenses; many are reported and are included in the data collection" (Pennsylvania Sentencing Commission, 1999)

[11] We are currently in the process of determining the precise randomization mechanism used in each of the six counties.

[12] Prior record score is a numeric variable calculated by PASC which aims to encapsulate the offender's entire prior criminal history.

felony drug convictions). At the charge level, PASC data allows observation of the offense classification for each charge and any mitigating or aggravating circumstances.

For the purposes of estimating the effect of incarceration, we then restrict our sample to the 6,127 offenders for whom we could locate valid correctional and arrest data. This restriction resulted in the removal of 282 (4.6%) offenders for whom either no rap sheet data could be located or for whom the correctional outcome was inconsistent with the sentencing data[13].

In this work, we measure reoffending by rearrest rate in 1, 2, 5, and 10 years after sentencing. To generate these rates, we use rearrest in the State of Pennsylvania as measured by Pennsylvania State Police rap sheet data. This rap sheet data allows us to observe any arrest that occurred in the state of Pennsylvania between the date of sentencing and April 30, 2010.

With respect to the calculation of our outcome measure, one point merits further discussion. Studies of the effect of imprisonment on reoffending, including analyses conducted by the authors (Nieuwbeerta, Nagin, and Blokland, 2009; Snodgrass, et al., in press) routinely correct for expose time—time not incarcerated—in calculating rearrest rates or time to rearrest. The rationale for the exposure time correction is to avert contamination of the behavioral effect of incarceration on reoffending with incapacitation effects.

In this analysis we do not correct for exposure time. Our changed stance on correcting for exposure time is reflective of several considerations. Because incarceration follows randomization, incarceration should be viewed as a consequence of treatment and as such should not be statistically controlled. Suppose, for example, incarceration exacerbates criminality. As a consequence, individuals who are initially incarcerated, on average, will commit more crimes and thereby, will be

---

[13] For example, sentencing data indicated a period of confinement in state prison but no period of confinement could be located in the Department of Corrections files.

more vulnerable to further stints of incarceration. Because their greater vulnerability to incarceration is a result of their treatment status, it should not be statistically controlled. Post-treatment assignment adjustments for exposure time, including for the initial "treatment-status" incarceration, also creates imbalances in age across treatment status. This is a very serious potential threat to identifying the treatment effect because recidivism is highly age dependent (Nagin, et al., 2009) with older adults offending at substantially lower rates than younger adults. By correcting for exposure time, incarcerated offenders are older and, hence, less likely to offend than those that are not incarcerated. This relationship between incarceration and offending, however, is a result of the aging that takes place during incarceration rather than "the effect" of incarceration[14]. Finally, not correcting for exposure time produces a treatment effect estimate that is more relevant from a policy perspective because it measures how many additional (fewer) offenses are incurred by society in the next *t* years due to the use of imprisonment.

METHODS

We use the instrumental variables approach advanced by Rosenbaum and colleagues (Imbens and Rosenbaum, 2005; Rosenbaum, 1996, 1999, 2002a, 2002b) to estimate the effect of imprisonment on 1 year, 2 year, 5 year, and 10 year rearrest rates. We begin with a general overview of the methodological approach and conclude with a more technical discussion.

We rely on the randomization of cases to judges within county as the basis for using judge as an instrument to identify the effect of incarceration. Like all instrumental variables techniques, this approach uses the variation in treatment induced by the instrument to identify the

---

[14] This also raises interesting questions about what exactly constitutes the treatment effect of incarceration. If one's conceptualization of the effect of incarceration includes the aging that takes place while incarcerated, then this concern is mitigated.

effect of treatment. In our application, this requires that the judge to whom the individual is randomized must impact the likelihood that an offender is incarcerated, net of the impact of other factors. The use of judge as an instrument also requires that the judge to whom an individual is randomized impacts the likelihood of rearrest only through his/her effect on the likelihood of incarceration. These requirements are sometimes referred to as an exclusion restriction. These two requirements play a central role in both the classical approaches to instrumental variables found in the econometric literature (e.g., Angrist, Imben, and Rubin, 1996) and the approach used here.

The first requirement, that the instrument induces variation in treatment net of the impact of other factors, can be resolved empirically. In this application, we demonstrate that this requirement is satisfied by examining differences across judges in their tendency to use incarceration as a punitive sanction. We demonstrate that such variation exists across judges in section 5.2. This approach suffices to demonstrate treatment variation across judges due to the properties of randomization. In particular, randomization guarantees that case and offender characteristics, whether measured or **unmeasured**, are equivalent across judges in a county. Hence, any variation in the use of incarceration must be attributable to a judge effect.

The second exclusion restriction requirement is an assumption that can be argued, but not empirically verified. In our application, the exclusion restriction requires that judges have no impact on the likelihood of recidivism beyond their impact on the likelihood of incarceration. This assumption is quite reasonable given the very limited interaction between judge and offender in most circumstances. However, it is possible that this condition could be violated. For example, a stern admonishment from the bench may deter (or exacerbate) future criminality in a

12

subset of offenders[15]. Alternatively, an informal request by the judge that local law enforcement more closely watch a given offender, thereby increasing the likelihood that s/he is observed engaging in criminal activity, would constitute a violation of the exclusion restriction. While these scenarios are possibilities, conversations with criminal justice practitioners in Pennsylvania indicate that they rarely occur.

As discussed above, this work uses the judge to whom a case is randomized as an instrument to identify the effect of incarceration. Our identification approach also rests on an important additional assumption –the effect of incarceration is homogeneous[16] and additive. These two assumptions play a critical role in the estimation strategy described below. To see why, consider the extreme case where one judge incarcerates his/her entire caseload and another judge incarcerates none of his/her caseload. If the treatment effect is additive and homogenous this implies that in expectation the difference in the rearrest rate of their respective caseloads will be β, where β measures the difference in the recidivism rate in a sanction regime in which all individuals are incarcerated versus a sanction regime in which none are incarcerated[17]. Because the treatment effect is assumed homogenous this implies that in the less extreme case where the difference in the probability of imprisonment between the harsh and lenient judge is Δ, in expectation the difference in the rearrest rate of their caseloads will be $\Delta\beta$.

The approach taken here uses the following line search to generate a $100(1-\propto)\%$ confidence set for the parameter β. The analyst first proposes a minimum value for β. For those that were incarcerated, this proposed value is subtracted from their observed rearrest rate, call

---

[15] However, respect for authority is not traditionally a hallmark of the criminally involved.
[16] This is a strong assumption, but one that can be relaxed. If their exists treatment effect heterogeneity, but judges do not sort offenders into prison based on the offender's return from prison, then this approach estimates the mean of the treatment effect distribution. See, for example, Heckman, Schmierer, and Urzua (2010).
[17] Under the assumptions used here, β can also be interpreted as the individual-level effect of incarceration.

this the adjusted rearrest rate[18]. Next, a test for the equality of mean adjusted rearrest rate is conducted across all judges in the county. If the test concludes that there is statistically significant variation across judge in mean adjusted rearrest rate, then this proposed value of β is rejected. The test is then repeated for a larger value of β until a value of β is found that generates mean equivalence in adjusted rearrest rate across judges. This value of β forms the lower bound of the $100(1-\alpha)\%$ confidence set of β. This testing process continues for successively larger values of β until the largest β that generates mean equivalence in adjusted rearrest rate is found. This largest β forms the upper bound of the confidence set.   If the set contains zero, then this implies that β cannot be signed.

To formalize the previous discussion, the *i*-th offender is randomly assigned to the *j*-th judge in county k. As previously discussed, we use six counties in this analysis, so $k \in \{Centre, Crawford, Cumberland, Dauphin, Erie, Mercer\}$. Henceforth we suppress the notation of k. Z_i is the judge to which the *i*-th offender is randomly assigned. We assume that the number of judges in a given county is fixed at z, so $Z_i \in \{1,2,\dots,z\}$ The *i*-th offender will be sentenced by the randomly assigned judge to either incarceration, $d_i = 1$, or released to the community, $d_i = 0$. This individual has a fixed potential response to treatment—the individual would be rearrested at rate $r_{I,i}$ if sentenced to prison (i.e., $d_i = 1$) and rate $r_{C,i}$ if released to the community (i.e., $d_i = 0$). We model individual response to incarceration as,

$$r_{I,i} - r_{C,i} = \beta d_i$$

---

[18] Since the effect is, by assumption, additive, and if the proposed value is plausible, the adjusted rearrest rate is the rate at which the incarcerated offender would have be rearrested if s/he had not been exposed to incarceration. Furthermore, the expected rate at which offenders would have been rearrested if not exposed to incarceration is balanced across judges due to randomization.

However, since $d_i = 1$ or $d_i = 0$, but never both, we only observe only one element of the of the pair $(r_{I,i}, r_{C,i})$. Namely, the realized rearrest rate, $R_i = d_i r_{I,1} + (1 - d_i)r_{C,i} = r_{C,i} + \beta d_i$.

Because β is assumed to be constant across i, if its value was known the adjusted rearrrest rate would be $R_i - \beta d_i = d_i r_{I,1} + (1 - d_i)r_{C,i} - \beta d_i = r_{C,i}$. The adjusted rearrest rate would not depend on whether the individual was incarcerated. Hence, in expectation, the adjusted rearrest rate would be invariant across judges who make differential use of incarceration. We do not, however, know β. To estimate β, the line search algorithm described above is used. To see why this approach is valid, suppose the proposal value is β_0, where $\beta_0 \neq \beta$. Then the adjusted rearrest rate is $R_i - \beta_0 d_i = r_{C,i} + \beta d_i - \beta_0 d_i = r_{C,i} + (\beta - \beta_0)d_i$. Thus, the adjusted rearrest rate would continue to depend on whether the individual was incarcerated. Consequently, the adjusted rearrest rate would vary across judges who make differential use of incarceration[19]. If the variation across judge in adjusted rearrest rate is statistically significant, then β_0 is not a plausible estimate of β.

As noted above, in order to generate our confidence intervals we combine the observed judge effect with a model. Put differently, the way that we interpret the observed judge effect is driven by our model, and our model may be incorrect. Our model assumes a constant, additive treatment effect. If the effect of incarceration varies across offenders, this may pose a significant problem for the approach taken here. How consequential such a violation would be depends on whether judges can discern the distribution of individual-level treatment effects. If judges cannot distinguish the offender-level effect of incarceration, then the approach outlined above estimates the mean of the distribution of incarceration effects. However, if the effect of incarceration varies across offenders and judges can discern the offender-level return from incarceration and

---

[19] Recall the even though β is fixed across i, $r_{C,i}$ varies. Thus, this relationship only holds in expectation.

this information is used to guide the sentencing decision, then our interpretation of the observed

judge effect as the average treatment effect no longer holds. See Manski and Nagin (1998) for a

demonstration of how such judge discernment capability can substantially affect the bound on

the treatment effect of incarceration. The effect crucially depends upon how the judge

incorporates this information into the sentencing decision.

It should also be noted that this analysis considers only the effect of the in/out decision –

it does not consider the impact of time served. That is, our model assumes that the dose-

response relationship between time served and rearrest rate is invariant with respect to time

served. If the effect of incarceration varies as a function of time served in a meaningful way,

then our estimate represents the dose-response function integrated out with respect to the density

of time served. Our model also does not consider the context of incarceration. It assumes the

effect of incarceration in a state prison is the identical to the effect of incarceration in a county

jail, and it further assumes that the conditions of confinement (e.g., security level of the facility,

distance from friend and family) do not influence the impact of incarceration. Although the

current work ignores these considerations for the sake of clarity and tractability, they are

important considerations.

RESULTS

RANDOMIZATION

As stated above, our methodological approach relies heavily on the randomization of cases to

judges. It is randomization which guarantees that the judge to whom a case is assigned is unrelated

to either the characteristics of offender or the offense. Without this property, the methodological

approach taken in this work would not suffice to identify any casual effect. To see why, suppose

two judge violated randomization by employing a trading scheme in which one judge exchanged a small number of serious cases for a large number of less serious cases with the second judge. If the tendency to recidivate is greater for those accused of serious offenses, then we would expect the judge receiving the more serious offenses to have greater average recidivism. However, this difference in recidivism would be attributable to differences in case load characteristics of the judges rather than the effect of incarceration. If the characteristics defining the swapped cases were measured, such a trading scheme would be detectable by comparing the distribution of offense characteristics between the two judges. For this reason we carefully check for differences in observed case characteristics across judges to test whether, at least based on measured covariates, randomization appeared to have successfully achieved balance. On the whole, there is little evidence against the randomization hypothesis in the six counties used in this analysis.

However, even with flawless adherence to a valid randomization procedure, substantively important covariate imbalance may persist after a single randomization. Moreover, a hypothesis test at the significance level α will incorrectly reject the null hypothesis of balance with probability α. We observe 42 different measurable characteristics related to offender demographics, current offense severity, and extent of prior criminal offending[20]. Many of these characteristics have been repeatedly found in the criminological literature to be predictive of the sentencing decision and of recidivism. We, thus, expect to find imbalance in about 4 measured covariates simply by chance. In half of our counties there is more balance than we would expect to observe by chance. In the remainder, we observe one or two more covariates out of balance than one would expect. Table 1 through Table 6 show the 42 covariates used in this study, the mean level of each covariate by judge,

_____

[20] To check covariate balance we use data for all 6,515 offenders sentenced in the six counties during 1999. We do so since randomization occurs at docketing and sample restrictions were made on the basis of outcomes that occur after both randomization and sentencing.

and whether there existed a statistically significant difference across judge. Table 7 shows the number of covariates out of balance in each county.

****

Insert Tables 1-6. Descriptive Statistics by Judge and County

****

****

Insert Table 7. Number of Covariates Out of Balance

****

We identified more imbalance than expected in three counties. However, the substantive size of the observed imbalance is small. Concerns of space and brevity prohibit a full discussion of each out of balance covariate in each county, but for the purpose of illustration we take a careful look at the imbalance detected in Crawford county. In Crawford county we found inter-judge variation in 6 measured covariates: the previous number of burglary adjudications, the previous number of aggravated assault conviction, the previous number of rape conviction, the maximum offense gravity score for the charges contained in the case, the average offense gravity score for the charges contained in the case, and the minimum offense gravity score for the charges contained in the case. This is two more covariates out of balance than we would expect to observe by chance.

On closer inspection, although statistically significant, the magnitude of the imbalance observed in Crawford county is small and likely of little substantive import. In 1999, two judges sentenced offenders in Crawford county, Judge M and Judge V. Judge M sentenced 300 cases while Judge V sentenced 269. Of the 300 cases sentenced by Judge M, two contained an offender with a

prior burglary adjudication. Of the 269 cases sentenced by Judge V, seven contained an offender with a prior burglary adjudication. Similarly, Judge M sentenced no offenders with either a prior rape conviction or a prior aggravated assault conviction. Judge V sentenced three offenders with a prior rape conviction and three offenders with a prior aggravated assault conviction. Thus, while we were able to detect a mean difference, the size of the difference is rather small.

The final three measures that differed across the two judges in Crawford county were the seriousness of the least severe charge in a case, the average seriousness of the charges in the case, and the seriousness of the most severe charge in a case[21]. To examine imbalance in these three measures, a kernel density smoother was estimated for each judge for each measure. This approach allows the direct inspection of an estimate of the distribution and does not rely solely on mean comparisons. These kernel density estimates are shown in Figures 2 through 4. Although using statistical tests we were able to identify a statistically significant mean difference in these measures between the two judges, Figure 2 through 4 shows that this mean difference was an artifact of small differences in the right tails of distribution. Indeed, examining other descriptive shows how similar the two distributions are. For all three seriousness covariates, the judge-specific distributions have the same 1st quartile, the same median, and the same 3rd quartile.

****

Insert Figure 2., 3., 4. Kernel Density Estimates

****

VARIATION IN INCARCERATION

---

[21] Where seriousness is defined as Offense Gravity Score (OGS).

As discussed in the methods section, in order to use judge as an instrument to estimate the effect of incarceration, we must first demonstrate that there exists substantial variation across judges in their willingness to use incarceration as a punitive sanction. Since randomization is conducted at the county level, all analyses are conducted within county. The blue bars in Figures 5 through 10 show the proportion of offenders sentenced to a period of incarceration by each judge. As shown in Table 2, there is statistically significant evidence of variation in the use of incarceration in five of the six counties used in this analysis.

****

Insert Figures 5 to 10 Variation in Confinement and Rearrest About Here

****

****

Insert Table 8 Variation in the use of Incarceration About Here

****

Centre was the lone county in which we could not detect differences in the use of confinement across judge. The harshest judge in Centre county incarcerated 39.4% of offenders s/he sentenced, while the most lenient judge incarcerated 37.5% of offenders s/he sentenced. The difference in incarceration rates observed across the three judges was not statistically significant (*p-value*=0.932). In the remaining five counties, there were large, statistically significant differences across judges in their use of incarceration. In Crawford county, Judge M incarcerated 59.4% of the sentenced, while Judge V incarcerated 70.9% of the sentence. This difference was statistically significant at traditional levels (*p-value*=0.008). In Cumberland county, the most lenient judge,

Judge B, incarcerated 46.6% of those s/he sentenced, while most punitive judge, Judge O, incarcerated 64.5% of those sentenced. The difference observed across the 5 judges was statistically significant (*p-value*<0.001). Variation in the use of incarceration was more dramatic in Dauphin county where the most lenient judge incarcerated one-fifth of offenders while the harshest judge incarcerated nearly two-thirds of those s/he sentenced (65.3%). The variation across the seven judges in Dauphin county was statistically significant at any traditional level (*p-value*<0.001). In Erie county, the difference across judges in the use of confinement was again statistically significant (*p-value*<0.001) with the harshest judge incarcerating 68.8% of the sentenced and the most lenient judge incarcerating 43.0% of those sentenced. Finally, in Mercer county, the observed difference across judge in the use of confinement was again statistically significant (*p-value*<0.001). The least punitive judge, Judge W, incarcerated 37.6% of those s/he sentenced, while the harshest judge, Judge F, incarcerated over three-quarters (76.5%) of those s/he sentenced.

ESTIMATED EFFECT OF INCARCERATION

Having demonstrated both covariate balance that is consistent with randomization and substantial inter-judge variation in the use of confinement, we now examine evidence of the effect of incarceration on 1 year, 2 year, 5 year, and 10 year rearrest rates. To do so, we apply the model developed in Section 4 to each of the six counties used in this work for each of the four outcomes. Again, the main result from this approach is a 95% confidence interval for the estimated effect of incarceration. Hence, we estimate twenty four 95% confidence interval. These are shown in Table 3. If the interval falls wholly below zero, then incarceration reduces subsequent criminality. If the interval falls entirely above zero, then incarceration exacerbates subsequent criminality. If the interval contains zero, then we cannot sign the effect of incarceration. No significant variation across judge in average rearrest rate implies 0 will be contained in our confidence interval. Put

differently, if we are unable to detect variation across judge in average rearrest rate, then our data does not sign the effect of incarceration. To see why, observe that no variation across judge in rearrest rate despite large differences in the use of confinement means $\Delta\beta = 0$ despite observing $\Delta \neq 0$. This clearly implies that there is insufficient evidence in the data to conclude that $\beta \neq 0$.

$$*********************************$$

Insert Table 9 County-Specific C.I. Estimates About Here

$$*********************************$$

In Centre county, there was no statistically significant evidence of variation across judges in their willingness to use incarceration as a punitive sanction. Put differently, the instrument is very weak in Centre county. Consequently, our ability to detect an effect in Centre county is seriously compromised. An adequate model should indicate this, and our does. As shown in Table 3, our confidence interval for the effect of incarceration in Centre county is simply $(-\infty, \infty)$. In essence, no variation in the use of confinement means that $\Delta = 0$, so $\Delta\beta = 0$ for any value of β. Stated differently any value of β is consistent with the data.

In the remaining five counties, there was a clear judge effect. Therefore, our instrument should aid in the estimation of the effect of incarceration. In Crawford county, despite differences in the willingness to use incarceration, there is little variation across judges in the average rearrest rate of their caseloads. Not surprisingly then, all confidence intervals contains zero. There is no evidence that incarceration impacts the rate at which offenders would be rearrested in the next year, the next two years, the next five years, or the next ten years. In Crawford county, our results indicate that, with high probability, exposure to incarceration could increase the rate at which

22

offenders are arrested in the next ten years by as much 0.18 arrests per year or could decrease the rate at which offenders are rearrested in the next ten years by up to 0.48 arrests per year.

Similarly, in Cumberland county there was again statistically significant evidence of a judge effect, but relatively little variation across judge in average rearrest rates. Based on the point estimates shown in Figure 7, average rearrest rates in the first year are slightly lower for judges who incarcerated a greater share of the offender they sentenced. This pattern, however, is no longer evident by the tenth year after sentencing. Thus, in Cumberland county, point estimates are consistent with a mild incapacitation effect. However, all confidence intervals again contain zero. Consistent with the point estimates, our confidence interval for the effect on one year rearrest rates, (-1.48, 0.18), is more heavily weighted toward a suppressing effect of incarceration. However, even in the first year after sentencing we are unable to distinguish the effect from 0. For the ten year window, ambiguity concerning the sign of the effect persists with the 95% CI covering the interval -0.42 to 0.23.

In Dauphin county, there was large and statistically significant variation across judge in the use of incarceration. Despite this, there is very little variation across judges in the rearrest rates of their caseloads. Not surprisingly then, all confidence intervals for the estimated effect of incarceration includes zero. For the 1 year window after sentencing, at the upper bound, incarceration is estimated to increase rearrest by as much as 0.22 arrests. However, at the lower bound, incarceration is estimated to decrease rearrest by 0.28 arrests. Similarly, for the ten year window the 95% confidence interval, ranging between -0.05 and 0.10, provides no evidence for signing the effect of incarceration on rearrest rates.

There is little evidence that incarceration impacts rates of rearrest in Erie county. In Figure 9, there is no clear association between rates of confinement and rates of rearrrest. This holds independent of the duration over which rearrest is measured. As would be expected given this, all confidence intervals contain zero. The estimated impact of exposure to incarceration on one year rearrest rates ranges between a decrease of 0.30 arrests and an increase of 0.18 arrests. Similarly, when examining the impact of incarceration on ten year rearrest rates, estimates range from a decrease of 0.10 arrests per year to an increase of 0.17 arrests per year.

Finally, in Mercer county, we again were able to distinguish a strong judge effect, but were unable to identify statistically significant variation across judge in rates of rearrest. Similar to what was observed in Cumberland county, based on point estimates there is weak evidence of a suppressing effect on rates of rearrest. Judges who incarcerated a greater share of offenders had slightly lower average rearrest rates over the observation periods considered here. Despite point estimates indicating a very mild suppressing effect, all confidence intervals for the effect of incarceration contain zero.

COUNTY-POOLED ESTIMATE OF THE EFFECT OF INCARCERATION

The confidence intervals thus far presented have implicitly allowed the effect of incarceration to vary freely across county. Although the notation has been suppressed, we have allowed β to be indexed on k. This approach allows for a great degree of flexibility since each county's confidence interval is generated without appealing to data from other counties. While quite flexible, this estimation strategy suffers from an important limitation—degradation of statistical power.

Statistical power has important scientific consequences for this analysis. When statistical power is low, even an otherwise well designed study may fail to detect a large effect. In this study, concerns of statistical power can be reframed as concerns about the width of the previously presented confidence intervals. As was noted above, each of the county-specific confidence intervals contain zero. That is, they do not allow us to either sign the effect of incarceration on recidivism or to determine if the effect is small whatever its sign. However, it could be the case that if the number of offenders sentenced in each county in 1999 was increased the confidence intervals would narrow sufficiently to exclude zero, thereby signing the effect of incarceration.

Although increasing the number of offenders in 1999 is not feasible, we can increase power by carefully pooling data across counties. To do so, we make an additional assumption – the effect of incarceration is the same across all six counties used in this analysis. To see why such an assumption allows us to pool data, we briefly revisit the model developed above.

Henceforth, we no longer suppress the notation on county. At the true, but unknown, value of β_k, the adjusted rate of rearrest of individual i sentenced in county k is $R_{i,k} - \beta_k d_{i,k} = r_{C,i,k}$. In other words, the adjusted rearrest rate is simply the rate at which the offender would have been rearrested if not exposed to incarceration. This rate is a fixed property of the individual, and hence, is balanced across judge by randomization. Put differently, $E[R_{i,k} - \beta_k d_{i,k} | K, Z] = E[r_{C,i,k} | K, Z]$ is balanced across Z within K at the true value of β_k. To generate our county-specific confidence intervals we use the previously defined line search to check if $E[R_{i,k} - \beta_{0,k} d_{i,k} | K, Z]$ is balanced across Z within K. The confidence interval is the set of $\beta_{0,k}$ that generate balance.

We can re-express this in a conceptually equivalent way through the use of a fixed effect linear model. Suppose we have a fixed effect linear model of the following form.

$$R_{i,k} - \beta_k d_{i,k} = \alpha + \sum_{i=2}^{z} \gamma_Z I(Z = i) + \varepsilon_{i,k}$$

Without loss of generality, assume that we omit judge 1 (i.e., Z=1) in county k. In this linear model, $\alpha = E[R_{i,k} - \beta_k d_{i,k}|K, Z = 1] = E[r_{C,i,k}|K, Z = 1] = E[r_{C,i,k}|K]$. In other words, at the true value of β_k, α captures the average adjusted rearrest rate for offenders in county k, which is simply the average rearrest rate for offenders in county k if none had been exposed to incarceration. Also note that $\gamma_Z = E[R_{i,k} - \beta_k d_{i,k}|K, Z = z] - E[R_{i,k} - \beta_k d_{i,k}|K, Z = 1] = E[r_{C,i,k}|K, Z = z] - E[r_{C,i,k}|K, Z = 1] = E[r_{C,i,k}|K] - E[r_{C,i,k}|K] = 0$. Stated less formally, γ_Z measures how the averge adjusted rearrest rate differs between the z-th judge and the omitted judge. Randomization implies that this is balanced at the true value of β_k. Hence, $\gamma_Z = 0 \; for \; Z = 2, ..., z \; in \; K$ at the true value of β_k. The fixed effect model can then be used to generate a confidence set by conducting the line search previously defined but substituting a test of the null $\gamma_Z = 0 \; for \; Z = 2, ..., z \; in \; K$ in the fixed effect model for a test of mean equality using ANOVA. The two approaches will generate virtually identical confidence intervals.

The added value of the fixed effects conceptualization is that can be immediately extended to accommodate multiple counties, thereby allowing the data to be pooled. Again, we layer on the additional assumption that the effect of incarceration does not vary across county. When including multiple counties, the model will take the following form,

$$R_{i,k} - \beta d_{i,k} = \alpha_{Centre} + \alpha_{Crawford} + \alpha_{Cumberland} + \alpha_{Dauphin} + \alpha_{Erie} + \alpha_{Mercer}$$

$$+ \sum_{i=2}^{z} \gamma_{Z,Centre}\, I_{Centre}(Z = i) + \sum_{i=2}^{z} \gamma_{Z,Crawford}\, I_{Crawford}(Z = i)$$

$$+ \sum_{i=2}^{z} \gamma_{Z,Cumberland}\, I_{Cumberland}(Z = i) + \sum_{i=2}^{z} \gamma_{Z,Dauphin}\, I_{Dauphin}(Z = i)$$

$$+ \sum_{i=2}^{z} \gamma_{Z,Erie}\, I_{Erie}(Z = i) + \sum_{i=2}^{z} \gamma_{Z,Mercer}\, I_{Mercer}(Z = i) + \varepsilon_{i,k}$$

As before, at the true value of β, each county-specific α measures the average rearrest rate of the offenders in the county had they not been exposed to incarceration. Given that there are likely idiosyncratic county-level factors that inform the tendency to recidivate (e.g., differing access to social services, differing labor markets and employment opportunities, differing distributions of socio-economic status), there is good reason to believe that the county-specific αs will not all be equal. This is not problematic. Also as before, the $\gamma_{Z,k}$ measure the difference in the average adjusted rearrest rate between judge Z and the omitted judge. At the true value of β, $\gamma_{Z,k} = 0\ for\ Z = 2, \dots, z\,, and\ all\ K$. To generate our joint confidence intervals which make use of all data, we use the line search previously outlined but test the null

$\gamma_{Z,k} = 0 for\ Z = 2, \dots, z\,, and\ all\ K$ in the fixed effects linear model. If we fail to reject this null hypothesis, then the proposal value is retained in the confidence interval.

****

Insert Table 10 Pooled C.I. Estimates About Here

****

Table 10 presents the cross-county pooled confidence intervals for each of the four post-release observations windows. As would be expected given our gain in statistical power and the nature of the estimation strategy, the cross-county confidence intervals are substantially tighter than those generated using only information from a single county. Similar to the findings from the county-specific confidence intervals, zero is contained in the 1year, 2year, and 10 year pooled confidence intervals. Put differently, when considering the effect of incarceration on recidivism there is again no evidence that incarceration impacts the rate at which offenders will be rearrested in the next 1, 2, or 10 years. With high probability, exposure to incarceration reduces the number of rearrests by no more than one-quarter of an arrest and increases the number of rearrests by no more than one-eighth of an arrest in the year after sentencing. Similarly, our pooled confidence interval indicates that in the two years after sentencing, incarceration reduces the number of yearly rearrests by no more than one-sixth of an arrest and increases the number of yearly arrests by no more than one-eighth of an arrest. When looking at the confidence interval employing the longest observation window, exposure to incarceration changes the yearly rate of rearrest by less than one-tenth of an arrest.

While these intervals are substantially narrower than those found in the county-specific analysis, they are still relatively wide when compared to the average rearrest rate. The average rearrest rate, across all six counties, in the first year after sentencing was 0.21 arrests, while our confidence interval was 0.36 arrests wide. In the first two years after sentencing, offenders sentenced in the six counties considered here average 0.19 rearrests per year, while our confidence interval for the estimated effect of incarceration is 0.27 arrests per year wide. Similarly, in the first ten years after sentencing, the average rearrest rate observed in the six counties used in this analysis was 0.17 arrests per year, while our confidence intervals had a

width of 0.15 rearrests per year. Put differently, pooling data substantially narrowed our confidence intervals, but the effect is still imprecisely estimated when measured against the tendency to recidivate.

The confidence interval examining the effect of incarceration on the tendency to recidivate in the five years after release merits further discussion. We could find no real value that was consistent with our data, thus producing an empty confidence set for the effect of incarceration in the five years after sentencing. There is more than one interesting explanation for this. First note that, given our estimation strategy, we should not be surprised that most of our pooled confidence intervals are essentially the intersection of the county-specific confidence intervals. When looking at the county-specific confidence intervals for the effect of incarceration on five year recidivism, the intersection of the county-specific confidence intervals is quite small, (0.00, 0.02). Given this small intersection and the improved statistical power, it is not at all surprising that no value consistent with the data could be found.

Alternatively, this could be that an indication that our assumption of a common treatment effect across counties is untenable. Although the evidence is quite fragile, this finding may indicate that the effect of incarceration depends on the community from which the offender is drawn. Just as it would not be surprising if the tendency to offend varied by county, it might not be surprising to find that the effect of incarceration varies across county. Again, however, the empirical support in this analysis for such conjecture is very weak.

DISCUSSION AND CONCLUSIONS

On the whole, the results provide no indication of whether the experience of incarceration increases or decreases reoffending rate. This holds whether we observe rearrest during a very short

window (one year) or a long window (ten years). The result holds across all six counties and persists even after pooling data to increase power. This result is quite consistent with an emerging body of work that uses randomization as the basis for concluding that incarceration has no clear impact on recidivism. Our results are very similar to Green and Winick (2010), Berube and Green (2007), and Loeffler (2011) who exploit the randomization of cases to judges to make a determination of the effect of incarceration on future offending. Similarly, using the same data analyzed here, Anwar and Stephens (2011) find no evidence that duration of confinement impacts criminality[22]. All arrive at the same conclusion – the data are not informative about even the sign of the effect. Further, compared to the base rearrest rate, the confidence intervals on the effect size in our analysis are generally sufficiently wide that it is not particularly informative about whether the effect size is small whether positive or negative.

We earlier argued why we believe this analysis advances the earlier work based on the randomization of cases to judges. In particular, rather than relying on standard IV regressions, we develop an individual-level model of the effect of incarceration and show how this model allows us to trace out the impact of incarceration. Importantly, the approach applied in this paper makes it clear when the instrument is weak and provides results that are reflective of the fact that with a weak instrument it will be difficult to precisely identify a treatment effect.

Still, our study suffers from several important limitations. First, like all IV analyses, it is model based. Thus, the validity of our results depends on the tenability of our model. As was discussed in detail in Section 4, our model assumes we have a valid instrument and that the effect of incarceration is constant and additive. We further assume that, at the individual level, the decision to

---

[22] Anwar and Stephens (2011), however, use a different methodological approach to estimate the dose-response function.

incarcerate is unrelated to the effect of incarceration on offending[23]. While we conclude that the

assumptions needed to use judge as an instrument are likely met in this application, we also believe

that the assumption of an additive and constant treatment effect is more fragile. It would thus be

valuable to examine the sensitivity of our conclusions to alternative formulations such as those posed

in Manski and Nagin (1998) that assume that judges can discern to some degree individual-level

response to incarceration and acts upon that knowledge in sentencing decisions.

One way that our model did incorporate the possibility of effect heterogeneity was through

the impact of county. Because we estimated the effect of incarceration independently in each of the

six counties, the effect could vary freely across county. Despite this flexibility, we were unable to

differentiate the effect from zero in any county. Notwithstanding, many arguments concerning the

effect of incarceration are built on the interaction between offender and community[24]. If there is

variation in the way that communities respond to offenders, there is likely to be variation in the

effect of incarceration. If the goal is the development of sound public policy, then understanding the

dynamic that exists between communities and those released from prison is an interesting and

important avenue for future work.

While allowing the effect to vary across county afforded flexibility, it also degraded power.

As in all empirical work, power plays a central role in the capability to distinguish the effect of

incarceration. This work used a sample of 6,127 offenders. However, when analyses were

conducted at the county level, sample sizes ranged from 501 in Centre county to 1,988 in Dauphin

county. These cases were randomized within county to between two judges (in Crawford county)

and seven judges (in Dauphin county). To confront this power issue, data was pooled to estimate

---

[23] This will be trivially true if our assumption of a constant treatment effect is true.
[24] For example, stigmatization by the community, ability to obtain gainful employment, building and maintaining pro-social relationships.

cross-county confidence intervals. Despite the improved power, we remained unable to uncover a relationship between incarceration and the tendency to recidivate.

It should also be noted that in this study we observe convictions, not cases. Randomization takes place when a case is docketed. Thus, judges have the opportunity to filter cases prior to our ability to observe them. Our concerns about this limitation are assuaged for two primary reasons. First, based on measurable characteristics, there is little difference across judges in the types of cases that progress to conviction. Second, based on conversations with court officials, rates of conviction in these counties are quite high, often above 90%. Thus, we fail to observe only a relatively small proportion of cases.

On the whole, the literature on the effect of incarceration is developing rapidly both in size and sophistication. This work aids in both of these respects, while also echoing the conclusions of the modern literature. There is little persuasive evidence that incarceration reduces future criminality.

REFERENCES

Angrist, Joshua, Guido W. Imbens, and Donald B. Rubin. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91:444-72.

Bergman, Gerald R. 1976. *The Evaluation of an Experimental Program Designed to Reduce Recidivism among Second Felony Criminal Offenders*. PhD Dissertation, Wayne State University, Department of Measurement and Evaluation.

Bernburg, Jon Gunnar, Marvin D. Krohn. 2003. Labeling, life chances, and adult crime: the direct and indirect effects of official intervention in adolescence on crime in early adulthood. *Criminology* 41:1287-1318.

Berube, Danton, and Donald P. Green. 2007. The effects of sentencing on recidivism: Results from a natural experiment." Unpublished Paper. New Haven, CT: Yale University.

Bound, Josh, David A. Jaeger, and Regina M. Baker. 1995. Problems with instrumental variables estimation when the correlation between the instrument and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association* 90:443-50.

Cullen, Francis T. 2002. Rehabilitation and treatment programs. In *Crime: Public Policies for Crime Control*, eds. James Q. Wilson and Joan Petersilia. Oakland, CA: ICS Press.

Freeman, Richard B. 1995. Why do so many young American men commit crimes and what might we do about it? *Journal of Economic Perspectives* 10:25-42.

Glueck, Sheldon, and Eleanor Glueck. 1950. *Unraveling delinquency*. New York: The Commonwealth Fund.

Green, Donald P., and Daniel Winik. 2010. Using random judge assignments to estimate the effects of incarceration and probation on recidivism among drug offenders. *Criminology* 48:357-87.

Hagan, John, and Alberto Palloni. 1990. The social reproduction of a criminal class in working-class London 1950-1980. *American Journal of Sociology* 96:265-99.

Hawkins, Gordon J. 1976. *The Prison: Policy and Practice*. Chicago: Chicago University Press.

Heckman, James J., Daniel Schmierer, and Sergio Urzua. 2010. Testing the correlated random coefficient model. *Journal of Econometrics* 158:177-203.

Imbens, Guido W., and Paul R. Rosenbaum. 2005. Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Association, Soc. A* 168:109-26.

Killias, Martin, Marcelo Aebi, and Denis Ribeaud. 2000. Does community service rehabilitate better than short-term imprisonment? Results of a controlled experiment. *Howard Journal* 39:40-57.

Klepper, Steven, and Daniel Nagin. 1989. The deterrent effect of perceived certainty and severity revisited. *Criminology* 27:721-46.

Laub, John H., and Robert J. Sampson. 2003. *Shared Beginnings, Divergent Lives: Delinquent Boys to Age 70.* Cambridge, MA: Harvard University Press.

Loeffler, Charles E. 2011. *Estimating the Effects of Imprisonment on the Life-Course.* Ph.D. Dissertation, Harvard University.

Layton MacKenzie, Doris. 2002. Reducing the criminal activities of known offenders and delinquents: Crime prevention in the courts and corrections. In L.W. Sherman, D.P. Farrington, B.C. Welsh, & D. Layton MacKenzie (Eds.), *Evidence-Based Crime Prevention*. London: Routledge.

Maddala, G.S., and Jinook Jeong. 1992. On the exact small sample distribution of the instrumental variable estimator. *Econometrica* 60:181-83.

Manski, Charles F., and Daniel Nagin. 1998. Bounding disagreements about treatment effects: a case study of sentencing and recidivism. *Sociological Methodology* 28:99-138.

Matsueda, Ross L. 1992. Reflected appraisal, parental labeling, and delinquency: Specifying a symbolic interactionist theory. *American Journal of Sociology* 97:1577-611.

Nagin, Daniel S., Francis T. Cullen, and Cheryl Lero Jonson. 2009. Imprisonment and re-offending. In M. Tonry (Ed.), *Crime and justice: A review of research*, *vol. 38*. Chicago: University of Chicago Press.

Nagin, Daniel S., and Raymond Paternoster. 1994. Personal capital and social control: The deterrence implications of a theory of individual differences in criminal offending. *Criminology* 32: 581-606.

Nagin, Daniel S., and Greg Pogarsky. 2003. An experimental investigation of deterrence: Cheating, self-serving bias, and impulsivity. *Criminology* 41:167-94.

Nagin, Daniel S., and Joel Waldfogel. 1995. The effects of criminality and conviction on the labor market status of young British offenders. *International Review of Law and Economics* 15:109-26.

Nagin, Daniel S., and Joel Waldfogel. 1998. The effect of conviction on income through the life cycle. *International Review of Law and Economics* 18:25-40.

Nelson, Charles R., and Richard Startz. 1990. Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica* 58:967-76.

Owens, Emily G. (2009). More time, less crime? Estimating the incapacitative effect of sentence enhancements. *Journal of law and Economics* 52:551-79.

Rosenbaum, Paul R. 1996. Comment on "Identification of causal effects using instrumental variables" by Angrist, Imbens, and Rubin. *Journal of the American Statistical Association* 91:465-68.

Rosenbaum, Paul R. 1999. Using quantile averages in matched observational studies. *Applied Statistics* 48:63-78.

Rosenbaum, Paul R. 2002a. *Observational Studies, 2nd Ed.* New York: Springer.

Rosenbaum, Paul R. 2002b. Covariance adjustment in randomized experiments and observational studies (with discussion). *Statistical Science* 17:286-327.

Sampson, Robert J., and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points Through Life.* Cambridge, MA: Harvard University Press.

Sampson, Robert J., and John H. Laub. 1997. A life-course theory of cumulative disadvantage and the stability of delinquency. In *Developmental Theories of Crime and Delinquency*, ed Terrence P. Thornberry. New Brunswick, NJ: Transaction Publishers.

Sherman, Lawrence W. 1992. *Policing and Domestic Violence.* New York: Free Press.

Steffensmeier, Darrell, and Jeffery T. Ulmer. 2005. *Confessions of a Dying Thief*. New Brunswick, NJ: Aldine/Transaction Publishers.

Sweeten, Gary, and Robert Apel. 2007. Incapacitation: Revisiting an old question with a new method and new data. *Journal of Quantitative Criminology* 23:303-26.

Van der Werff, C. (1979). *Speciale Preventie*. Den Haag (NL): WODC.

Waldfogel, Joel. 1994. The effect of criminal conviction on income and the trust reposed in the workmen. *Journal of Human Resources* 29:62-81.

Western, Bruce. 2002. The impact of incarceration on wage mobility and inequality. *American Sociological Review* 67:526-46.

Western, Bruce, Jeffery R. Kling, and David F. Weiman. 2001. The labor market consequences of incarceration. *Crime and Delinquency* 47:410-27.

Williams, Kirk R., and Richard Hawkins. 1986. Perceptual research on general deterrence: A critical overview. *Law and Society Review* 20:545-72.
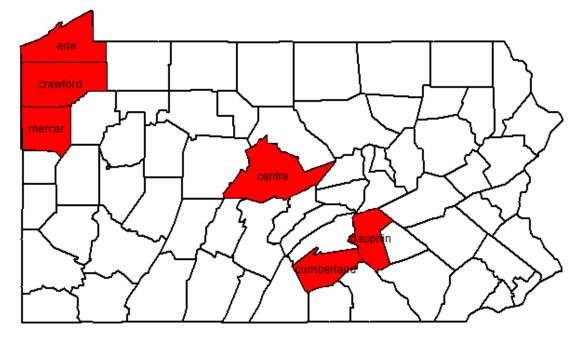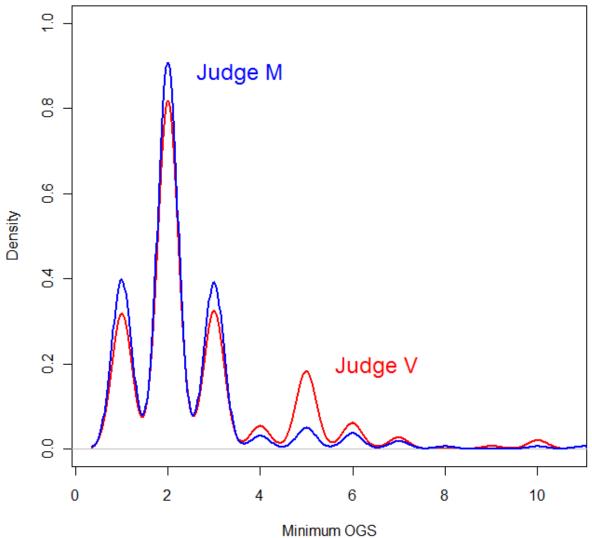
Figure 1. Map of Pennsylvania, by County

Counties Used In Analysis Shown in Red

Figure 2. Kernel Density Smoother of Offense Gravity Score of Least Severe Charge in Case

Figure 3. Kernel Density Smoother of Mean Offense Gravity Score in Case

Figure 4. Kernel Density Smoother of Offense Gravity Score of Most Severe Charge in Case

Maximum OGS
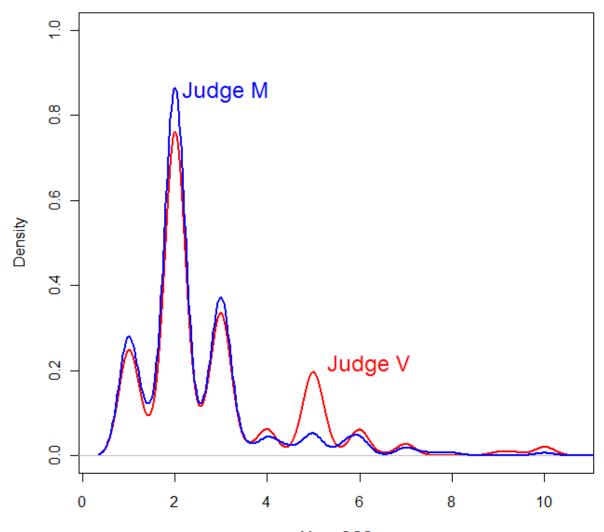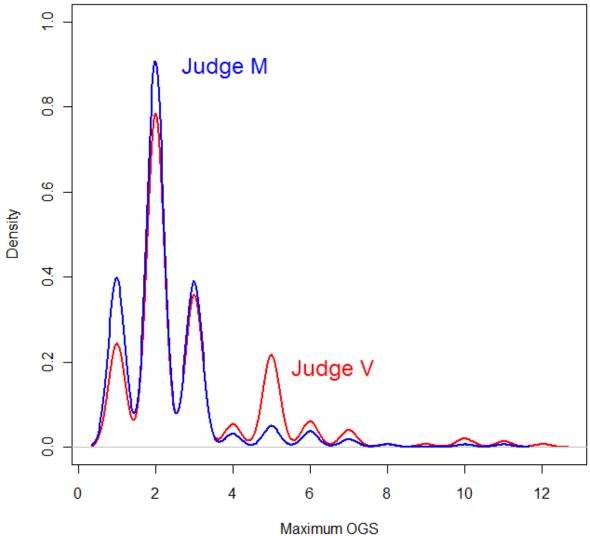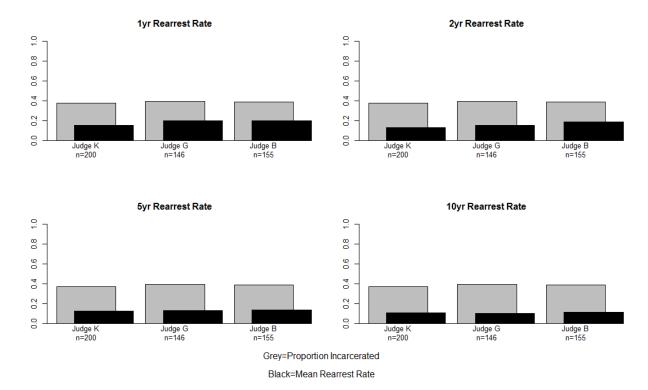Both Judges have Identical 1st, 2nd, and 3rd Quartiles

Figure 5. Centre County – Incarceration and Average Rearrest Rate

1yr Rearrest Rate

Judge K
n=200

Judge G
n=146

Judge B
n=155

2yr Rearrest Rate

Judge K
n=200

Judge G
n=146

Judge B
n=155

5yr Rearrest Rate

Judge K
n=200

Judge G
n=146

Judge B
n=155

10yr Rearrest Rate

Judge K
n=200

Judge G
n=146

Judge B
n=155

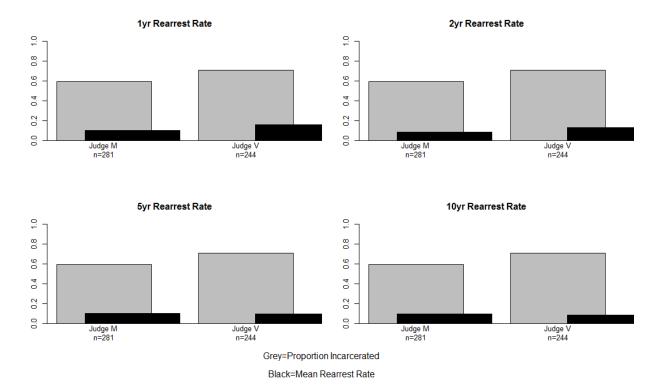Grey=Proportion Incarcerated

Black=Mean Rearrest Rate

Figure 6. Crawford County – Incarceration and Average Rearrest Rate

Grey=Proportion Incarcerated
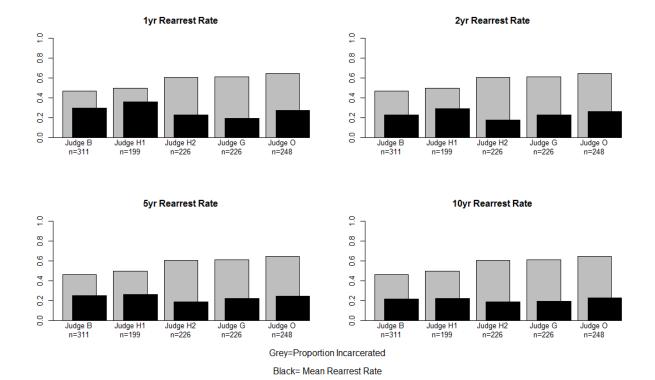
Black=Mean Rearrest Rate

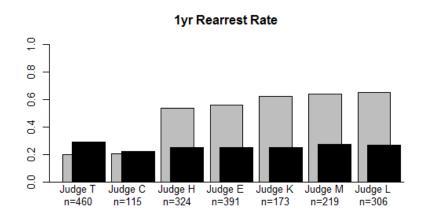## Figure 7. Cumberland County – Incarceration and Average Rearrest Rate

Grey=Proportion Incarcerated

Black= Mean Rearrest Rate

Figure 8. Dauphin County – Incarceration and Average Rearrest Rate

Grey=Proportion Incarcerated

Black=Mean Rearrest Rate
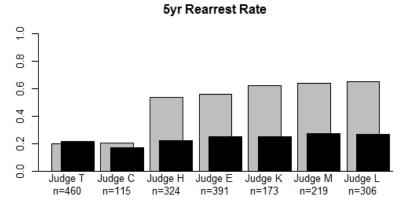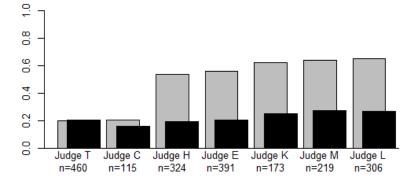
Figure 9. Erie County – Incarceration and Average Rearrest Rate
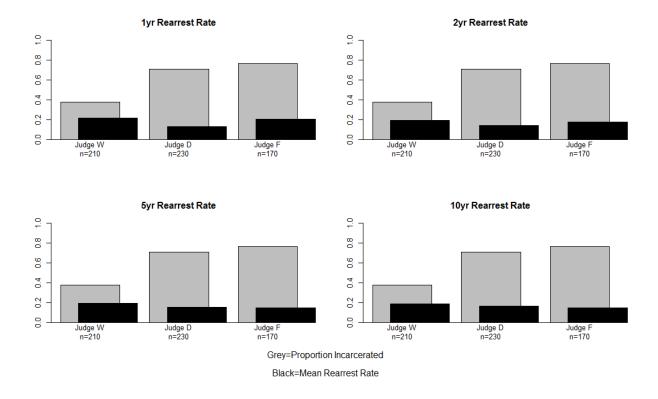

Grey=Proportion Incarcerated

Black=Mean Rearrest Rate

Figure 10. Mercer County – Incarceration and Average Rearrest Rate

Grey=Proportion Incarcerated

Black=Mean Rearrest Rate

Table 1. Descriptive Statistics for Convictions, by County and Judge – Centre County

	Judge K n_{random}=211 $n_{analysis}$=200		Judge G n_{random}=159 $n_{analysis}$=146		Judge B n_{random}=163 $n_{analysis}$=155		Sig.
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	
Offender Demographics							
Male	0.839	0.368	0.849	0.358	0.859	0.348	
Race – White	0.905	0.293	0.899	0.300	0.896	0.306	
Race – Black	0.090	0.286	0.063	0.243	0.092	0.289	
Race – Hispanic	0.000	0.000	0.019	0.136	0.000	0.000	*
Race – Asian	0.000	0.000	0.006	0.079	0.012	0.110	
Race – American Indian	0.005	0.069	0.000	0.000	0.000	0.000	
Race – Unknown or Other	0.000	0.000	0.012	0.012	0.000	0.000	
Age at Sentencing	30.45	10.10	32.05	11.78	29.08	10.31	*
Current Offense Characteristics							
Seriousness of Least Severe	2.51	1.54	2.50	1.56	2.68	1.75	
Average Seriousness	2.74	1.64	2.78	1.73	2.97	1.80	
Seriousness of Most Severe	2.97	1.93	3.05	2.05	3.24	2.09	
Age at Offense (Average)	30.22	10.16	31.76	11.79	28.87	10.27	.
Deadly Weapon Enhancement	0.009	0.094	0.019	0.137	0.000	0.000	
Prior Juvenile Adjudications							
Murder	0.000	0.000	0.000	0.000	0.000	0.000	
Voluntary Manslaughter	0.000	0.000	0.000	0.000	0.000	0.000	
Aggravated Assault	0.000	0.000	0.006	0.080	0.000	0.000	
Rape	0.000	0.000	0.000	0.000	0.012	0.157	
IDSI	0.000	0.000	0.000	0.000	0.000	0.000	
Kidnapping	0.000	0.000	0.000	0.000	0.000	0.000	
Arson	0.000	0.000	0.000	0.000	0.000	0.000	
Robbery	0.000	0.000	0.000	0.000	0.000	0.000	
Burglary	0.005	0.069	0.006	0.080	0.006	0.078	
Other Felony I&II	0.000	0.000	0.019	0.018	0.000	0.000	
Felony Drug	0.000	0.000	0.000	0.000	0.000	0.000	
Felony III	0.000	0.000	0.019	0.177	0.006	0.078	
Prior Adult Conviction							
Murder	0.000	0.000	0.000	0.000	0.000	0.000	
Voluntary Manslaughter	0.000	0.000	0.000	0.000	0.000	0.000	
Aggravated Assault	0.009	0.097	0.019	0.177	0.018	0.135	
Rape	0.005	0.069	0.006	0.080	0.006	0.078	
IDSI	0.000	0.000	0.000	0.000	0.000	0.000	
Kidnapping	0.000	0.000	0.000	0.000	0.000	0.000	
Arson	0.000	0.000	0.006	0.080	0.000	0.000	
Robbery	0.000	0.000	0.006	0.080	0.012	0.110	
Burglary	0.014	0.119	0.165	1.526	0.031	0.206	
Other Felony I&II	0.019	0.168	0.038	0.222	0.025	0.247	
Felony Drug	0.019	0.217	0.063	0.501	0.012	0.110	
Felony III	0.076	0.451	0.209	1.211	0.190	0.850	
Other Misdemeanor	0.828	1.748	1.196	3.383	0.665	1.654	
Overall Measure of Prior Record Severity (PRS)							
Min PRS	0.512	1.189	0.767	1.548	0.558	1.479	

Mean PRS	0.512	1.189	0.779	1.547	0.558	1.479
Max PRS	0.512	1.189	0.792	1.555	0.558	1.479
PRS varies?	0.000	0.000	0.013	0.111	0.000	0.000
For Those in Randomization						
Incarcerated	0.370	0.483	0.377	0.484	0.409	0.492
Minimum Duration (months)	3.28	11.83	6.19	37.21	3.47	12.70
Maximum Duration (months)	11.80	24.67	20.04	93.81	11.73	25.86
Sentenced to Life	0.006	0.077	0.000	0.000	0.005	0.071
For Those in Analysis						
Incarcerated	0.375	0.484	0.390	0.488	0.394	0.489
Minimum Duration (months)	3.39	12.13	6.45	38.81	3.65	13.00
Maximum Duration (months)	12.10	25.20	20.84	97.85	12.35	26.39
Sentenced to Life	0.006	0.077	0.000	0.000	0.005	0.071

. – 0.05< p-value < 0.1
* – 0.01< p-value < 0.05
** – 0.001<p-value< 0.01
*** – p-value<0.0001

For continuous covariates, p-values reflect One-Way ANOVA test for equality of mean. For binary covariates, p-values reflect a chi-square test of the equality of proportions. n_{random} is the number of offenders sentenced by the judge who were used in the randomization checks. $n_{analysis}$ is the number of offenders sentenced by the judge who were used in the analysis estimating the effect of incarceration,

Table 2. Descriptive Statistics for Convictions, by County and Judge – Crawford County

	Judge M n_{random}=300 $n_{analysis}$=281		Judge V n_{random}=269 $n_{analysis}$=244		Sig.
	$\bar{x}$	s	$\bar{x}$	s	
Offender Demographics					
Male	0.833	0.373	0.822	0.383	
Race – White	0.950	0.218	0.944	0.229	
Race – Black	0.047	0.211	0.056	0.229	
Race – Hispanic	0.000	0.000	0.000	0.000	
Race – Asian	0.000	0.000	0.000	0.000	
Race – American Indian	0.000	0.000	0.000	0.000	
Race – Unknown or Other	0.003	0.058	0.000	0.000	
Age at Sentencing	33.60	11.65	32.74	11.75	
Current Offense Characteristics					
Seriousness of Least Severe	2.319	1.349	2.690	1.650	**
Average Seriousness	2.498	1.452	2.828	1.690	*
Seriousness of Most Severe	2.698	1.774	2.970	1.925	.
Age at Offense (Average)	33.10	11.63	32.30	11.66	
Deadly Weapon Enhancement	0.017	0.128	0.004	0.063	
Prior Juvenile Adjudications					
Murder	0.000	0.000	0.000	0.000	
Voluntary Manslaughter	0.000	0.000	0.000	0.000	
Aggravated Assault	0.000	0.000	0.004	0.061	
Rape	0.000	0.000	0.000	0.000	
IDSI	0.000	0.000	0.000	0.000	
Kidnapping	0.000	0.000	0.000	0.000	
Arson	0.000	0.000	0.004	0.061	
Robbery	0.000	0.000	0.007	0.086	
Burglary	0.007	0.082	0.034	0.235	.
Other Felony I&II	0.007	0.082	0.019	0.136	
Felony Drug	0.003	0.058	0.004	0.061	
Felony III	0.010	0.100	0.019	0.136	
Prior Adult Conviction					
Murder	0.000	0.000	0.000	0.000	
Voluntary Manslaughter	0.000	0.000	0.000	0.000	
Aggravated Assault	0.000	0.000	0.011	0.106	.
Rape	0.000	0.000	0.011	0.106	.
IDSI	0.000	0.000	0.000	0.000	
Kidnapping	0.000	0.000	0.000	0.000	
Arson	0.000	0.000	0.000	0.000	
Robbery	0.013	0.182	0.000	0.000	
Burglary	0.047	0.241	0.082	0.389	
Other Felony I&II	0.087	0.469	0.060	0.253	
Felony Drug	0.027	0.181	0.052	0.240	
Felony III	0.137	0.460	0.116	0.623	
Other Misdemeanor	1.124	1.640	0.944	1.425	
Overall Measure of Prior Record Severity (PRS)					
Min PRS	0.873	1.494	1.033	1.631	

Mean PRS	0.881	1.495	1.035	1.630	
Max PRS	0.883	1.498	1.037	1.630	
PRS varies?	0.003	0.058	0.004	0.061	
For Those in Randomization					
Incarcerated	0.590	0.492	0.691	0.462	.
Minimum Duration (months)	3.66	17.1	4.15	10.79	
Maximum Duration (months)	182.3	372.3	173.0	357.4	
Sentenced to Life	0.000	0.000	0.000	0.000	
For Those in Analysis					
Incarcerated	0.594	0.491	0.709	0.454	**
Minimum Duration (months)	3.79	17.96	4.20	10.79	
Maximum Duration (months)	190.6	378.9	164.2	349.2	
Sentenced to Life	0.000	0.000	0.000	0.000	

. – 0.05< p-value < 0.1
* – 0.01< p-value < 0.05
** – 0.001<p-value< 0.01
*** – p-value<0.0001

For continuous covariates, p-values reflect One-Way ANOVA test for equality of mean. For binary covariates, p-values reflect a chi-square test of the equality of proportions. n_{random} is the number of offenders sentenced by the judge who were used in the randomization checks. $n_{analysis}$ is the number of offenders sentenced by the judge who were used in the analysis estimating the effect of incarceration,

Table 3. Descriptive Statistics for Convictions, by County and Judge – Cumberland County

	Judge B $n_{random}=317$ $n_{analysis}=311$		Judge H1 $n_{random}=209$ $n_{analysis}=199$		Judge H2 $n_{random}=238$ $n_{analysis}=226$		Judge G $n_{random}=235$ $n_{analysis}=226$		Judge O $n_{random}=258$ $n_{analysis}=248$		Sig.
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	
Offender Demographics											
Male	0.800	0.400	0.837	0.370	0.828	0.378	0.868	0.339	0.790	0.407	
Race – White	0.845	0.361	0.842	0.365	0.878	0.327	0.851	0.356	0.872	0.334	
Race – Black	0.136	0.342	0.144	0.351	0.105	0.307	0.128	0.334	0.116	0.321	
Race – Hispanic	0.009	0.097	0.000	0.000	0.008	0.091	0.013	0.112	0.008	0.088	
Race – Asian	0.003	0.056	0.010	0.010	0.008	0.091	0.004	0.065	0.000	0.000	
Race – American Indian	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Race – Unknown or Other	0.006	0.077	0.005	0.071	0.000	0.000	0.004	0.063	0.004	0.063	
Age at Sentencing	31.77	10.54	30.84	10.48	33.42	10.34	32.20	10.78	32.34	10.42	
Current Offense Characteristics											
Seriousness of Least Severe	2.561	1.586	2.714	1.629	2.885	2.130	2.983	2.082	2.518	1.403	*
Average Seriousness	2.786	1.658	2.887	1.775	3.093	2.176	3.264	2.243	2.669	1.420	**
Seriousness of Most Severe	3.016	1.964	3.058	2.062	3.328	2.474	3.547	2.629	2.839	1.653	**
Age at Offense (Average)	31.39	10.41	30.58	10.34	32.99	10.25	31.91	10.84	32.03	10.39	
Deadly Weapon Enhancement	0.009	0.009	0.005	0.071	0.013	0.113	0.009	0.009	0.000	0.000	
Prior Juvenile Adjudications											
Murder	0.000	0.000	0.000	0.000	0.004	0.065	0.000	0.000	0.000	0.000	
Voluntary Manslaughter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Aggravated Assault	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Rape	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
IDSI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Kidnapping	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Arson	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Robbery	0.003	0.056	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Burglary	0.003	0.056	0.010	0.138	0.000	0.000	0.004	0.065	0.000	0.000	
Other Felony I&II	0.000	0.000	0.010	0.138	0.000	0.000	0.000	0.000	0.008	0.088	
Felony Drug	0.000	0.000	0.005	0.069	0.000	0.000	0.008	0.130	0.000	0.000	
Felony III	0.000	0.000	0.005	0.069	0.008	0.130	0.009	0.092	0.000	0.000	
Prior Adult Conviction											
Murder	0.003	0.056	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	

Voluntary Manslaughter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Aggravated Assault	0.016	0.125	0.019	0.137	0.013	0.112	0.013	0.113	0.047	0.350	
Rape	0.000	0.000	0.005	0.069	0.000	0.000	0.000	0.000	0.004	0.062	
IDSI	0.000	0.000	0.005	0.069	0.000	0.000	0.000	0.000	0.000	0.000	
Kidnapping	0.010	0.097	0.005	0.069	0.000	0.000	0.000	0.000	0.000	0.000	
Arson	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Robbery	0.019	0.177	0.043	0.384	0.030	0.170	0.000	0.000	0.031	0.214	
Burglary	0.086	0.394	0.062	0.279	0.101	0.366	0.051	0.221	0.128	0.486	*
Other Felony I&II	0.073	0.316	0.038	0.216	0.051	0.256	0.119	0.456	0.047	0.245	
Felony Drug	0.159	0.505	0.124	0.409	0.131	0.407	0.140	0.524	0.116	0.397	
Felony III	0.165	0.671	0.234	0.706	0.195	0.649	0.174	0.514	0.252	0.843	
Other Misdemeanor	1.312	1.854	1.385	1.928	1.521	2.692	1.200	1.664	1.240	2.068	
Overall Measure of Prior Record Severity (PRS)											
Min PRS	1.303	1.836	1.378	1.725	1.202	1.719	1.115	1.672	1.267	1.923	
Mean PRS	1.324	1.840	1.404	1.751	1.203	1.718	1.115	1.672	1.280	1.926	
Max PRS	1.341	1.862	1.416	1.782	1.210	1.718	1.115	1.672	1.287	1.936	
PRS varies?	0.009	0.097	0.010	0.097	0.004	0.065	0.000	0.000	0.004	0.062	
For Those in Randomization											
Incarcerated	0.486	0.500	0.498	0.500	0.605	0.489	0.600	0.490	0.640	0.480	***
Minimum Duration (months)	3.33	18.36	4.21	26.29	4.56	12.68	4.72	13.26	2.38	5.71	
Maximum Duration (months)	25.58	103.0	25.01	88.60	23.14	68.73	21.58	32.98	19.35	62.73	
Sentenced to Life	0.003	0.055	0.000	0.000	0.000	0.000	0.004	0.063	0.000	0.000	
For Those in Analysis											
Incarcerated	0.466	0.499	0.497	0.500	0.606	0.489	0.611	0.782	0.645	0.479	***
Minimum Duration (months)	3.20	18.36	3.78	26.19	4.76	12.98	4.75	13.44	2.32	5.68	
Maximum Duration (months)	25.46	103.9	24.11	88.66	19.43	26.77	21.70	33.28	19.57	63.91	
Sentenced to Life	0.003	0.055	0.000	0.000	0.000	0.000	0.004	0.063	0.000	0.000	

. – 0.05< p-value < 0.1
* – 0.01< p-value < 0.05
** – 0.001<p-value< 0.01
*** – p-value<0.0001

For continuous covariates, p-values reflect One-Way ANOVA test for equality of mean. For binary covariates, p-values reflect a chi-square test of the equality of proportions. n_{random} is the number of offenders sentenced by the judge who were used in the randomization checks. $n_{analysis}$ is the number of offenders sentenced by the judge who were used in the analysis estimating the effect of incarceration,

Table 4. Descriptive Statistics for Convictions, by County and Judge – Dauphin County

	Judge T n_{random}=471 $n_{analysis}$=460		Judge C n_{random}=118 $n_{analysis}$=115		Judge H n_{random}=336 $n_{analysis}$=324		Judge E n_{random}=402 $n_{analysis}$=391		Judge K n_{random}=180 $n_{analysis}$=173	
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
Offender Demographics										
Male	0.793	0.405	0.795	0.404	0.821	0.464	0.802	0.473	0.784	0.467
Race – White	0.503	0.500	0.364	0.481	0.464	0.499	0.473	0.499	0.467	0.499
Race – Black	0.427	0.495	0.576	0.494	0.438	0.496	0.430	0.495	0.467	0.499
Race – Hispanic	0.023	0.151	0.042	0.201	0.051	0.219	0.047	0.212	0.042	0.201
Race – Asian	0.004	0.065	0.000	0.000	0.015	0.121	0.005	0.070	0.000	0.000
Race – American Indian	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Race – Unknown or Other	0.042	0.202	0.017	0.129	0.033	0.178	0.045	0.207	0.033	0.180
Age at Sentencing	31.67	10.86	30.48	10.28	32.27	10.80	31.68	10.71	31.57	10.55
Current Offense Characteristics										
Seriousness of Least Severe	2.96	2.20	2.81	1.75	2.65	1.70	2.93	1.95	2.60	2.02
Average Seriousness	3.49	2.15	3.52	2.01	3.05	1.64	3.41	1.92	2.99	2.05
Seriousness of Most Severe	4.04	2.55	4.12	2.66	3.49	1.94	3.94	2.35	3.39	2.36
Age at Offense (Average)	31.11	10.76	30.10	10.08	31.68	10.55	31.25	10.71	31.14	10.62
Deadly Weapon Enhancement	0.021	0.144	0.076	0.265	0.015	0.121	0.017	0.131	0.011	0.105
Prior Juvenile Adjudications										
Murder	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.050	0.000	0.000
Voluntary Manslaughter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Aggravated Assault	0.011	0.104	0.000	0.000	0.009	0.095	0.010	0.123	0.022	0.149
Rape	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.050	0.000	0.000
IDSI	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.050	0.000	0.000
Kidnapping	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Arson	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Robbery	0.000	0.000	0.000	0.000	0.003	0.055	0.003	0.050	0.000	0.000
Burglary	0.007	0.081	0.000	0.000	0.009	0.123	0.020	0.174	0.011	0.106
Other Felony I&II	0.011	0.123	0.034	0.183	0.006	0.078	0.013	0.112	0.017	0.129
Felony Drug	0.026	0.207	0.026	0.159	0.036	0.203	0.038	0.217	0.000	0.000
Felony III	0.041	0.210	0.052	0.290	0.033	0.196	0.015	0.123	0.073	0.319

Prior Adult Conviction										
Murder	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Voluntary Manslaughter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Aggravated Assault	0.196	0.180	0.017	0.131	0.018	0.134	0.025	0.157	0.022	0.149
Rape	0.000	0.000	0.000	0.000	0.012	0.110	0.005	0.071	0.006	0.075
IDSI	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.050	0.000	0.000
Kidnapping	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.050	0.000	0.000
Arson	0.000	0.000	0.000	0.000	0.003	0.055	0.003	0.050	0.000	0.000
Robbery	0.039	0.269	0.026	0.159	0.042	0.267	0.086	0.624	0.039	0.195
Burglary	0.100	0.411	0.043	0.204	0.055	0.265	0.109	0.441	0.096	0.495
Other Felony I&II	0.054	0.271	0.060	0.273	0.067	0.250	0.056	0.261	0.084	0.317
Felony Drug	0.222	0.582	0.241	0.668	0.188	0.563	0.183	0.502	0.225	0.547
Felony III	0.237	0.985	0.147	0.594	0.170	0.482	0.173	0.597	0.129	0.521
Other Misdemeanor	1.031	1.743	1.060	1.706	1.076	1.487	1.227	1.745	0.944	1.356
Overall Measure of Prior Record Severity (PRS)										
Min PRS	1.42	1.95	1.42	1.86	1.39	1.81	1.62	2.02	1.42	1.85
Mean PRS	1.45	1.94	1.45	1.86	1.40	1.81	1.64	2.02	1.43	1.84
Max PRS	1.49	1.98	1.51	1.89	1.43	1.84	1.66	2.05	1.47	1.88
PRS varies?	0.017	0.129	0.025	0.157	0.012	0.108	0.010	0.099	0.011	0.105
For Those in Randomization Data										
Incarcerated	0.159	0.396	0.212	0.409	0.542	0.498	0.554	0.497	0.628	0.483
Minimum Duration (months)	3.57	10.16	5.95	17.36	4.17	8.97	5.40	13.80	6.10	12.92
Maximum Duration (months)	41.28	175.7	95.04	267.8	50.03	184.5	42.06	156.5	42.89	148.2
Sentenced to Life	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
For Those Used in Analysis										
Incarcerated	0.200	0.400	0.209	0.407	0.540	0.498	0.558	0.497	0.642	0.479
Minimum Duration (months)	3.63	10.25	5.58	17.26	3.93	7.82	5.35	13.87	5.76	12.56
Maximum Duration (months)	42.21	177.7	95.86	271.2	50.82	187.6	42.66	158.5	42.80	150.8
Sentenced to Life	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

. – 0.05< p-value < 0.1
* – 0.01< p-value < 0.05
** – 0.001<p-value< 0.01
*** – p-value<0.001

	Judge M $n_{random}=225$ $n_{analysis}=219$		Judge L $n_{random}=315$ $n_{analysis}=306$		Sig.
	$\bar{x}$	s	$\bar{x}$	s	
Offender Demographics					
Male	0.804	0.397	0.755	0.483	
Race – White	0.480	0.500	0.483	0.500	
Race – Black	0.471	0.499	0.454	0.498	
Race – Hispanic	0.031	0.174	0.006	0.079	*
Race – Asian	0.004	0.067	0.006	0.079	
Race – American Indian	0.000	0.000	0.000	0.000	
Race – Unknown or Other	0.013	0.115	0.051	0.220	
Age at Sentencing	31.39	10.42	31.12	10.42	
Current Offense Characteristics					
Seriousness of Least Severe	2.76	1.96	2.67	1.63	
Average Seriousness	3.22	1.96	3.08	1.64	**
Seriousness of Most Severe	3.69	2.43	3.49	2.09	***
Age at Offense (Average)	30.97	9.52	30.93	10.31	
Deadly Weapon Enhancement	0.009	0.094	0.029	0.167	**
Prior Juvenile Adjudications					
Murder	0.000	0.000	0.000	0.000	
Voluntary Manslaughter	0.000	0.000	0.000	0.000	
Aggravated Assault	0.004	0.067	0.000	0.000	
Rape	0.000	0.000	0.000	0.000	
IDSI	0.000	0.000	0.000	0.000	
Kidnapping	0.000	0.000	0.000	0.000	
Arson	0.000	0.000	0.000	0.000	
Robbery	0.000	0.000	0.003	0.057	
Burglary	0.009	0.134	0.003	0.057	
Other Felony I&II	0.022	0.148	0.010	0.098	
Felony Drug	0.004	0.067	0.026	0.179	
Felony III	0.148	0.175	0.211	0.224	.
Prior Adult Conviction					
Murder	0.004	0.067	0.000	0.000	
Voluntary Manslaughter	0.000	0.000	0.000	0.000	

Aggravated Assault	0.049	0.288	0.019	0.138	
Rape	0.009	0.094	0.000	0.000	
IDSI	0.000	0.000	0.000	0.000	
Kidnapping	0.000	0.000	0.000	0.000	
Arson	0.000	0.000	0.000	0.000	
Robbery	0.036	0.230	0.026	0.159	
Burglary	0.094	0.450	0.071	0.380	
Other Felony I&II	0.045	0.326	0.068	0.265	
Felony Drug	0.247	0.702	0.127	0.427	
Felony III	0.175	0.616	0.224	0.773	
Other Misdemeanor	1.199	1.670	1.118	1.899	

Overall Measure of Prior Record Severity (PRS)

Min PRS	1.34	1.89	1.23	1.73	
Mean PRS	1.34	1.89	1.23	1.73	
Max PRS	1.34	1.89	1.23	1.73	
PRS varies?	0.004	0.067	0.000	0.000	

For Those in Randomization

Incarcerated	0.636	0.481	0.660	0.474	***
Minimum Duration (months)	5.75	12.33	5.72	14.58	.
Maximum Duration (months)	32.95	133.3	26.64	85.50	*
Sentenced to Life	0.000	0.000	0.000	0.000	

For Those in Analysis

Incarcerated	0.639	0.780	0.654	0.472	***
Minimum Duration (months)	5.72	12.37	4.96	11.47	
Maximum Duration (months)	33.47	135.0	24.80	83.65	*
Sentenced to Life	0.000	0.000	0.000	0.000	

. – 0.05< p-value < 0.1
* – 0.01< p-value < 0.05
** – 0.001<p-value< 0.01
*** – p-value<0.001

For continuous covariates, p-values reflect One-Way ANOVA test for equality of mean. For binary covariates, p-values reflect a chi-square test of the equality of proportions. n_{random} is the number of offenders sentenced by the judge who were used in the randomization checks. $n_{analysis}$ is the number of offenders sentenced by the judge who were used in the analysis estimating the effect of incarceration,

Table 5. Descriptive Statistics for Convictions, by County and Judge – Erie County

	Judge B $n_{random}=242$ $n_{analysis}=235$		Judge D $n_{random}=481$ $n_{analysis}=463$		Judge C $n_{random}=399$ $n_{analysis}=374$		Judge A $n_{random}=233$ $n_{analysis}=221$		*Sig.*
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	S	
Offender Demographics									
Male	0.851	0.356	0.783	0.412	0.824	0.381	0.840	0.367	.
Race – White	0.740	0.439	0.705	0.456	0.652	0.476	0.639	0.480	*
Race – Black	0.223	0.416	0.222	0.416	0.296	0.456	0.283	0.451	*
Race – Hispanic	0.017	0.127	0.040	0.195	0.035	0.184	0.052	0.221	
Race – Asian	0.000	0.000	0.002	0.046	0.000	0.000	0.009	0.092	
Race – American Indian	0.000	0.000	0.002	0.046	0.000	0.000	0.009	0.092	
Race – Unknown or Other	0.021	0.142	0.031	0.174	0.015	0.122	0.009	0.092	
Age at Sentencing	30.97	10.43	32.19	10.91	30.42	10.28	30.14	10.09	*
Current Offense Characteristics									
Seriousness of Least Severe	2.606	1.538	2.583	1.699	2.398	1.639	2.716	1.904	
Average Seriousness	3.000	1.643	2.852	1.734	2.811	1.703	3.028	1.976	
Seriousness of Most Severe	3.465	2.179	3.149	1.996	3.270	2.205	3.422	2.373	
Age at Offense (Average)	30.61	10.39	31.58	10.82	29.95	10.35	29.55	10.18	*
Deadly Weapon Enhancement	0.033	0.179	0.040	0.195	0.033	0.178	0.052	0.221	
Prior Juvenile Adjudications									
Murder	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Voluntary Manslaughter	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.066	
Aggravated Assault	0.000	0.000	0.023	0.150	0.018	0.132	0.009	0.092	.
Rape	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
IDSI	0.000	0.000	0.000	0.000	0.003	0.050	0.000	0.000	
Kidnapping	0.000	0.000	0.002	0.046	0.000	0.000	0.000	0.000	
Arson	0.004	0.065	0.000	0.000	0.000	0.000	0.000	0.000	
Robbery	0.004	0.065	0.010	0.102	0.008	0.112	0.009	0.092	
Burglary	0.025	0.156	0.029	0.181	0.045	0.209	0.034	0.182	
Other Felony I&II	0.046	0.246	0.035	0.207	0.035	0.185	0.064	0.263	
Felony Drug	0.008	0.091	0.004	0.064	0.020	0.158	0.013	0.113	
Felony III	0.029	0.169	0.038	0.211	0.038	0.191	0.073	0.292	
Prior Adult Conviction									
Murder									
Voluntary Manslaughter									

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Aggravated Assault	0.000	0.000	0.002	0.046	0.003	0.050	0.004	0.066	
Rape	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
IDSI	0.004	0.065	0.008	0.091	0.003	0.050	0.000	0.000	
Kidnapping	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Arson	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Robbery	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.006	
Burglary	0.004	0.065	0.000	0.000	0.000	0.000	0.000	0.000	
Other Felony I&II	0.025	0.181	0.017	0.128	0.018	0.150	0.017	0.130	
Felony Drug	0.063	0.484	0.052	0.353	0.051	0.298	0.034	0.182	
Felony III	0.086	0.425	0.063	0.330	0.109	0.397	0.069	0.314	
Other Misdemeanor	0.063	0.275	0.050	0.270	0.066	0.311	0.077	0.363	
	0.188	0.692	0.119	0.549	0.152	0.525	0.124	0.497	
Overall Measure of Prior Record Severity (PRS)	0.950	1.782	1.074	2.142	1.013	1.678	1.021	2.116	
Min PRS									
Mean PRS									
Max PRS	1.136	1.731	1.085	1.683	1.271	1.856	1.146	1.678	
PRS varies?	1.140	1.730	1.088	1.682	1.278	1.827	1.156	1.674	
	1.145	1.731	1.091	1.682	1.286	1.840	1.167	1.677	
For Those in Randomization	0.004	0.064	0.004	0.064	0.003	0.050	0.013	0.113	
Incarcerated									
Minimum Duration (months)									
Maximum Duration (months)	0.430	0.495	0.482	0.500	0.627	0.484	0.700	0.458	***
Sentenced to Life	6.94	19.54	6.56	20.61	9.16	26.55	7.73	18.34	
	22.70	76.59	28.46	90.83	24.18	52.70	20.36	37.03	
For Those in Analysis	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Incarcerated									
Minimum Duration (months)									
Maximum Duration (months)	0.430	0.495	0.482	0.500	0.623	0.485	0.688	0.463	***
Sentenced to Life	6.43	18.06	6.72	20.98	8.77	25.96	7.92	18.80	
	21.59	74.79	29.07	92.47	23.37	51.26	20.64	37.97	
. − 0.05< p-value < 0.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
* − 0.01< p-value < 0.05									
** − 0.001<p-value< 0.01									
*** − p-value<0.0001									

For continuous covariates, p-values reflect One-Way ANOVA test for equality of mean. For binary covariates, p-values reflect a chi-square test of the equality of proportions. n_{random} is the number of offenders sentenced by the judge who were used in the randomization checks. $n_{analysis}$ is the number of offenders sentenced by the judge who were used in the analysis estimating the effect of incarceration,

Table 6. Descriptive Statistics for Convictions, by County and Judge – Mercer County

	Judge W $n_{random}=225$ $n_{analysis}=210$		Judge D $n_{random}=239$ $n_{analysis}=230$		Judge F $n_{random}=180$ $n_{analysis}=170$		Sig.
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	
Offender Demographics							
Male	0.771	0.420	0.860	0.347	0.810	0.392	*
Race – White	0.676	0.468	0.762	0.426	0.722	0.448	
Race – Black	0.240	0.427	0.218	0.413	0.222	0.416	
Race – Hispanic	0.000	0.000	0.000	0.000	0.000	0.000	
Race – Asian	0.000	0.000	0.008	0.091	0.000	0.000	
Race – American Indian	0.000	0.000	0.000	0.000	0.006	0.074	
Race – Unknown or Other	0.084	0.278	0.013	0.111	0.050	0.218	**
Age at Sentencing	32.76	10.26	32.94	10.27	33.71	10.30	
Current Offense Characteristics							
Seriousness of Least Severe	2.876	1.792	2.760	1.953	2.904	1.912	
Average Seriousness	2.992	1.820	2.890	1.970	3.062	1.934	
Seriousness of Most Severe	3.106	1.886	3.039	2.136	3.215	2.072	
Age at Offense (Average)	32.27	10.07	32.45	10.34	33.22	10.22	
Deadly Weapon Enhancement	0.000	0.000	0.008	0.091	0.011	0.105	
Prior Juvenile Adjudications							
Murder	0.000	0.000	0.000	0.000	0.000	0.000	
Voluntary Manslaughter	0.000	0.000	0.000	0.000	0.000	0.000	
Aggravated Assault	0.000	0.000	0.000	0.000	0.000	0.000	
Rape	0.000	0.000	0.004	0.065	0.000	0.000	
IDSI	0.000	0.000	0.004	0.065	0.000	0.000	
Kidnapping	0.000	0.000	0.000	0.000	0.000	0.000	
Arson	0.000	0.000	0.000	0.000	0.000	0.000	
Robbery	0.000	0.000	0.000	0.000	0.000	0.000	
Burglary	0.005	0.068	0.080	0.739	0.017	0.129	
Other Felony I&II	0.008	0.091	0.000	0.000	0.000	0.000	
Felony Drug	0.000	0.000	0.008	0091	0.000	0.000	
Felony III	0.014	0.151	0.017	0.158	0.011	0.106	
Prior Adult Conviction							
Murder	0.000	0.000	0.004	0.065	0.000	0.000	
Voluntary Manslaughter	0.000	0.000	0.000	0.000	0.000	0.000	
Aggravated Assault	0.005	0.068	0.008	0.130	0.006	0.075	
Rape	0.000	0.000	0.000	0.000	0.000	0.000	
IDSI	0.000	0.000	0.0000.	0.000	0.000	0.000	
Kidnapping	0.000	0.000	000	0.000	0.000	0.000	
Arson	0.000	0.000	0.004	0.065	0.000	0.000	
Robbery	0.018	0.134	0.021	0.144	0.006	0.075	
Burglary	0.046	0.267	0.038	0.231	0.040	0.390	
Other Felony I&II	0.084	0.341	0.084	0.358	0.051	0.287	
Felony Drug	0.027	0.164	0.013	0.112	0.023	0.149	
Felony III	0.068	0.345	0.092	0.356	0.169	0.670	.
Other Misdemeanor	1.023	2.040	0.785	1.168	1.234	2.588	.
Overall Measure of Prior Record Severity (PRS)							

Min PRS	0.733	1.433	0.921	1.611	0.800	1.481	
Mean PRS	0.753	1.440	0.921	1.611	0.806	1.480	
Max PRS	0.773	1.481	0.921	1.611	0.811	1.483	
PRS varies?	0.009	0.094	0.000	0.000	0.006	0.074	
For Those in Randomization							
Incarcerated	0.400	0.490	0.707	0.455	0.767	0.423	***
Minimum Duration (months)	2.80	6.87	4.62	11.45	5.42	12.39	*
Maximum Duration (months)	44.77	175.0	26.18	92.78	27.97	78.53	
Sentenced to Life							
For Those in Analysis							
Incarcerated	0.376	0.484	0.709	0.454	0.765	0.424	***
Minimum Duration (months)	2.54	6.76	4.64	11.59	5.49	12.62	*
Maximum Duration (months)	45.51	180.8	26.63	94.53	28.35	80.67	
Sentenced to Life	0.000	0.000	0.000	0.000	0.006	0.077	

. – 0.05< p-value < 0.1
* – 0.01< p-value < 0.05
** – 0.001<p-value< 0.01
*** – p-value<0.0001

For continuous covariates, p-values reflect One-Way ANOVA test for equality of mean. For binary covariates, p-values reflect a chi-square test of the equality of proportions. n_{random} is the number of offenders sentenced by the judge who were used in the randomization checks. $n_{analysis}$ is the number of offenders sentenced by the judge who were used in the analysis estimating the effect of incarceration,

Table7. Number of Out-of-Balance Covariates

County	Number
Centre	3
Crawford	4
Cumberland	6
Dauphin	5
Erie	6
Mercer	4

Note: Balance examined across 42 covariates. At the 0.1 level, we would expect to observe approximately 4 out of balance covariate.

Table 8. Variation in the Use of Incarceration

County	Least Punitive Judge		Most Punitive Judge		p-value
	% Inc.		**% Inc.**		
Centre	37.5	%	39.4	%	0.932
Crawford	59.4		70.9		0.008
Cumberland	46.6		64.5		<0.001
Dauphin	20.0		65.4		<0.001
Erie	43.0		68.8		<0.001
Mercer	37.6		76.5		<0.001

Note: p-values reflect a Fisher's exact (or chi-square) testing the equality of the proportion confined across all judges in the county.

Table 9. Estimated Impact of Incarceration on Rearrest Rate, by County

County	1 Year Rearrest Rate	2 Year Rearrest Rate	5 Year Rearrest Rate	10 Year Rearrest Rate
	95% C.I.	95% C.I.	95% C.I.	95% C.I.
Centre	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
Crawford	(-0.16, 2.49)	(-0.03, 1.82)	(-0.46, 0.30)	(-0.48, 0.18)
Cumberland	(-1.48, 0.18)	(-0.39, 0.27)	(-0.62, 0.14)	(-0.42, 0.23)
Dauphin	(-0.28, 0.22)	(-0.20, 0.17)	(0.00, 0.12)	(-0.05, 0.10)
Erie	(-0.30, 0.18)	(-0.22, 0.23)	(-0.16, 0.18)	(-0.10, 0.17)
Mercer	(-0.38, 0.11)	(-0.31, 0.12)	(-0.29, 0.02)	(-0.25, 0.06)

Table 10. Estimated Impact of Incarceration on Rearrest Rate, Pooled Estimate

County	1 Year Rearrest Rate	2 Year Rearrest Rate	5 Year Rearrest Rate	10 Year Rearrest Rate
	95% C.I.	95% C.I.	95% C.I.	95% C.I.
All	$(-0.24, 0.12)$	$(-0.15, 0.12)$	$\emptyset$	$(-0.08, 0.07)$