This is a preprint of an article published in [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(2): 357-387]. Pagination of this preprint may differ slightly from the published version.

USING RANDOM JUDGE ASSIGNMENTS TO ESTIMATE THE EFFECTS OF INCARCERATION AND PROBATION ON RECIDIVISM AMONG DRUG OFFENDERS*

DONALD P. GREEN
Department of Political Science
Yale University

DANIEL WINIK Yale
Law School

KEYWORDS: recidivism, specific deterrence, drug crime, natural experiments

Most prior studies of recidivism have used observational data to estimate the causal effect of imprisonment or probation on the probability that a convicted individual is rearrested after release. Few studies have taken advantage of the fact that, in some jurisdictions, defendants are assigned randomly to judges who vary in sentencing tendencies. This study investigates whether defendants who are assigned randomly to more punitive judges have different recidivism probabilities than defendants who are assigned to relatively lenient judges. We track 1,003 defendants charged with drug-related offenses who were assigned randomly to nine judicial calendars between June 1, 2002 and May 9, 2003. Judges on these calendars meted out sentences that varied substantially in terms of prison and probation time. We tracked defendants using court records across a 4-year period after the disposition of their cases

* The authors are grateful to Dan Cipullo, Debbie Grafton, Gregory Hale, and Nancy McKinney, who facilitated access to the District of Columbia Superior Court's public records. We thank Lara Chausow, Amy Plovnick, Josh Rosmarin, and Melissa Yuckel, who helped assemble the data, and Terence Leong and Shang Ha, who helped prepare the replication programs. Support for this project was provided by the Institution for Social and Policy Studies at Yale University, which bears no responsibility for the conclusions we draw. Direct correspondence to Donald P. Green, Department of Political Science, Yale University, 77 Prospect St., New Haven, CT 06520-8209 (e-mail: donald.green@yale.edu).

to determine whether they subsequently were rearrested. Our results indicate that randomly assigned variations in prison and probation time have no detectable effect on rates of rearrest. The findings suggest that, at least among those facing drug-related charges, incarceration and supervision seem not to deter subsequent criminal behavior.

A central and enduring question in the study of criminal behavior concerns the extent to which punishment diminishes a convict's likelihood of committing crimes in the future. The extensive empirical literature on recidivism has generated a range of different conclusions (Nagin, Cullen, and Jonson, 2009; Villettaz, Killias, and Zoder, 2006). Some studies suggest that those who are punished more severely become less likely to reoffend (Smith and Gartin, 1989); others contend that they become more likely to reoffend (Spohn and Holleran, 2002); and still others find no relationship between punishment and recidivism (Gottfredson, 1999). From this diverse array of empirical findings has sprung an equally diverse array of theories to explain why punishment makes criminals more attentive to the likelihood of arrest and the severity of punishment or, conversely, hardens criminals, brings them together with other criminals in ways that expand their criminal opportunities, and diminishes their ability to make a living through ordinary employment (for competing perspectives, see Blumstein, Cohen, and Nagin, 1978; Massoglia and Macmillan, 2002; Orsagh and Chen, 1988; Sherman, 1993).

Any or all of these theoretical accounts might be true, but the empirical foundation on which they are based is open to question. As Killias and Villettaz (2008) pointed out, the basic problem with the extant recidivism literature is that, with a few exceptions, it is based on observational data. Observational studies such as Smith and Gartin (1989) and Spohn and Holleran (2002), for example, track felons over time and use regression analysis to test whether, controlling for criminal history, those who are subjected to the harshest punishments are more or less likely to be rear-rested. As Achen (1986), Manski and Nagin (1998), and others have noted, this approach is susceptible to bias insofar as the unobserved attributes of a defendant that lead to harsher sentences also might affect the defendant's probability of rearrest.

An alternative approach makes use of the fact that judges are assigned at random to defendants in certain jurisdictions. To the extent that randomly assigned judges have different sentencing tendencies, a component of the sentences that defendants receive is a function of chance. For example, Martin, Annan, and Forst (1993) studied 367 drunk-driving defendants who were assigned randomly to one of two judges to estimate the effect of incarceration on subsequent arrest for drunk driving. A much larger study was Berube and Green's (2007) analysis of thousands of felony defendants who were assigned randomly to judges of varying punitiveness in the District of Columbia Superior Court between 1978 and 1984. The Berube and Green study had the advantage of tracking large numbers of felony defendants for more than a decade. Like Martin, Annan, and Forst (1993), Berube and Green found imprisonment to have little long-term deterrent effect. This finding is consistent with results from Killias, Aebi, and Ribeaud's (2000) study of 123 Swiss convicts who were assigned randomly to short prison terms or community service as well as with mixed results from two other small experiments and one natural experiment discussed in Villettaz, Killias, and Zoder (2006: 13–5).

Like the Berube and Green (2007) study, this study looks at defendants appearing before the District of Columbia Superior Court, but its focus is more specific in terms of time period and offender type, and it brings to bear more detailed information about both the prison and the probation sentence given to each defendant.[1] We restricted our attention to defendants facing solely drug-related charges. The growing scholarly interest in drug offenders (Wilson, Mitchell, and MacKenzie, 2006) reflects the unusual importance of drug-related incarceration in the American criminal justice system. Punishment of this kind of crime surged dramatically during and after the 1980s (Murakawa, 2005). More than 250,000 state prisoners (as of the end of 2005) and more than 95,000 federal prisoners (as of 2007) were behind bars on drug-related charges (West and Sabol, 2008). The extent to which incarceration affects recidivism among those convicted for drug-related offenses is of special relevance for ongoing policy debates and speaks to literatures in sociology, economics, and criminology on the responsiveness of criminal activity to sanctions (Persson and Siven, 2006; Robinson and Darley, 2004; Sherman, 1993; Sherman and Berk, 1984).

This article is organized as follows. We begin by describing the sample and the natural experiment from which our data are drawn. Second, we summarize the statistical requirements for an unbiased assessment of the causal relationship between sentencing and recidivism. Next, we describe patterns of sentencing and recidivism in our data. In particular, we show how the random assignment of judges leads to significant variations in the rate and duration of incarceration and probation sentences, which sets the stage for a test of whether exogenous variation in sentencing affects rates of recidivism. Our results suggest that longer sentences of incarceration do not diminish rates of recidivism. Probation periods likewise have no

1. Berube and Green (2007) lacked information about the prior criminal records of the defendants and about the length of time during which the prisoners were incarcerated. They did not examine the effects of probation, and their analysis of incarceration looked at the effects of the minimum sentence during a period of indeterminate sentences.

detectable effect on rates of rearrest. The concluding section discusses the policy implications of our findings and suggests directions for future research.

DATA AND MEASURES

SAMPLE

Our sample of defendants was gathered from public lockup lists and case file records from the District of Columbia Superior Court (hereafter, Court). Lockup lists are daily records of all criminal arrests in the District of Columbia and include the defendant's identifying information and the charged offenses. Information gathered from the lockup lists was supplemented by variables from the Court's electronic case-management database. The contents of this database are public record—a digital version of paper case files. The record for each case includes the defendant's name, date of birth, race, gender, address, and police and correctional identifiers. It lists additional or lesser charges introduced at various stages; the presiding judge at the time of sentencing; the dates of arrest, disposition, and sentencing; the final disposition; the sentence imposed (which might vary somewhat from the sentence actually served); and records of any postdisposition hearings on probation revocation. The database enables searches by name or identifier numbers for all the past and pending cases of a defendant in the District of Columbia. This function permitted us to record prior offenses and subsequent recidivism.[2] We also used Maryland's online court records database to observe whether defendants recidivated in that state.

We restricted our attention to defendants charged with felony drug offenses and no nondrug-related criminal offenses between June 1, 2002 and May 9, 2003.[3] Drug felonies in the District of Columbia comprise two primary offenses: distribution and possession with intent to distribute (PWID). Each offense can be qualified as "attempted" or "while armed," but such charges are rare and are excluded from the data set. We retained

2. The database includes records for all cases that are arraigned in the Court and not only those that proceed beyond arraignment. Our measure of recidivism, therefore, includes subsequent arrests in which the government decides not to prosecute (i.e., the case is "no papered").

3. Under the Sentencing Reform Amendment Act of 2000, the District of Columbia applies determinate sentences for offenses committed after August 5, 2000 (District of Columbia Sentencing Commission, 2000). Determinate sentences require offenders to serve at least 85 percent of their prison sentence because "good time credit" is limited to 15 percent [D.C. Code § 24-403.01(d); 18 U.S.C. § 3624(b)(1)]. The period of our study also predates the voluntary sentencing guidelines that took effect on June 14, 2004.

cases in which drug misdemeanors, typically the possession of small quantities of drugs or of drug paraphernalia, were charged along with the felony or felonies. We also retained a few cases in which minor quality-of-life misdemeanors (such as panhandling, possession of an open container of alcohol, public intoxication or urination, or disorderly conduct) were recorded on the lockup list as well as those that listed notifications of bench warrant or fugitive status.

We made two additional exclusions from the sample. First, we rejected a few cases in which the defendant had another case sentenced concurrently or disposed at the time of sentencing. Such dispositions—generally the result of global plea agreements—reflect a kind of "treatment" that spans multiple cases, charges, and judges and, therefore, cannot be analyzed reliably. Second, we excluded a few cases involving judges not on the normal calendars described in the following discussion.[4]

SENTENCING

Felony sentences in the District of Columbia include incarceration, probation, or both. When a sentence includes probation, it typically also includes a period of incarceration suspended on the condition that it will be imposed if the defendant fails to comply with the terms of probation. We recorded the total duration of incarceration imposed, the amount suspended, and the duration of probation, all in months. We defined nonsuspended incarceration (the time to be served regardless of the future conduct of the defendant) as the difference between total and suspended incarceration. We recorded the sentences as imposed, not as actually served, although the two in practice are similar. Defendants must serve probation in full, but as explained in footnote 3, they might serve from 85 percent to 100 percent of their prison term.

Incarceration and probation are two distinct aspects of sentence severity associated with different mechanisms of preventing recidivism. Nonsuspended incarceration is the immediate and tangible penalty to the defendant. Prison time can affect a defendant's propensity to recidivate both by incapacitating him and by causing him to recognize the price of crime, which deters him specifically from future criminal behavior. The second aspect of a sentence, which has been treated less fully in the literature, is ongoing supervision (Kim, 1994; Petersilia, 1997). When a defendant is

4. The reason for an abnormal judge assignment was not always clear, but one reason might be that cases are shifted on rare occasions, or "certed out," to senior judges or others not on the regular docket. These exclusions do not introduce bias because neither global plea agreements nor aberrant judges should be associated in any way (other than random variation) with the characteristics of particular cases or with the initial judge assignments.

released, either after a prison term or with no incarceration at all, he typically faces supervision by the Court Services and Offender Supervision Agency.[5] He must report for drug tests, avoid additional arrests, and so on. Probation time might influence recidivism rates in three ways. First, criminal behavior is more likely to be detected when the defendant is under official supervision. Second, a defendant on probation faces more stringent behavioral constraints, and so has more opportunities to violate the law, than a defendant not on probation. Third, a defendant on probation might commit additional crimes to defy the societal constraints being applied to him or her (Sherman, 1993).

Table 1 shows the bivariate distribution of incarceration and probation among the 1,003 defendants in our sample. We see that 272 (27 percent) of the 1,003 defendants received no punishment (of whom 242 had their cases dismissed before conviction or acquittal,[6] 18 were acquitted at trial, and 12 pleaded guilty and were sentenced to time served or to fully suspended sentences without probation). Another 312 (31 percent) received no prison time but were placed on probation.[7] The remaining 42 percent of our sample received some nonsuspended prison time. However, the average sentence was fairly brief. Approximately 56 percent of those who were sentenced to prison were incarcerated for 1 year or less, and another 29 percent received sentences of up to 2 years. Just 65 defendants received sentences of longer than 2 years; of this group, only 19 defendants were sentenced to serve 4 years or more. Incapacitation creates a built-in bias toward lower rates of rearrest among those receiving harsher prison terms, but this bias was small in our sample because of the short prison sentences imposed on most defendants. We monitored recidivism during a four-year period; therefore, fully 97.8 percent of our defendants had at least 1 year during which to recidivate, and 93.5 percent had at least 2 years.

Table 1 also conveys the frequency with which defendants were sentenced to probation. Excluding the 272 defendants who received no punishment leaves 731 defendants, 436 of whom were sentenced to probation instead of, or in addition to, prison time. Overall, probation tends to be correlated negatively with nonsuspended prison time. Of those sentenced

5. Defendants sentenced to more than 1 year in prison receive a mandatory minimum period of supervised release. We looked for evidence of a differential treatment effect for prison terms exceeding 1 year but found no evidence of it.

6. "Dismissed" cases are those in which charges initially were filed but later were dismissed by the court. In some cases, the government requested dismissal; in others, the judge dismissed the case because the government was unprepared to go forward; in others, the grand jury declined to return an indictment. This category of cases does not include those in which no charges ever were filed; these "no papered" cases were not assigned to a judicial calendar and, therefore, were excluded from our data set in the first place.

7. We use "prison time" to refer to *nonsuspended* prison sentences.

Table 1. Bivariate Distribution of Prison Sentence and Probation Sentence

Length of Probation Sentence	Length of Prison Sentence					Total
	Zero	0 < Prison ≤ 1 Year	1 < Prison ≤ 2 Years	2 < Prison ≤ 3 Years	Prison > 3 Years	
Zero years	272	138	101	36	20	567
0 < Probation ≤ 1 year	82	9	0	0	0	91
1 < Probation ≤ 2 years	168	65	12	6	2	253
2 < Probation ≤ 3 years	53	19	7	1	0	80
Probation > 3 years	9	3	0	0	0	12
Total	584	234	120	43	22	1,003

NOTES: Entries are the number of defendants receiving each combination of prison and probation sentence. The recidivism rates among those who were neither incarcerated nor put on probation was 56 percent (n = 152 [out of 272]). Those who were put on probation but not incarcerated had a 51 percent rate of recidivism (n = 160 [out of 312]) as compared with 47 percent among defendants who were incarcerated but not placed on probation (n = 140 [out of 295]). Those who both were incarcerated and put on probation recidivated at a rate of 61 percent (n = 76 [out of 124]).

to a nonzero prison term of 1 year or less, 41 percent received probation. Just 16 percent of those sentenced to between 1 and 2 years in prison received probation terms.

RECIDIVISM

Tracking recidivism introduces a range of conceptual and measurement issues, as many authors have noted (Blumstein and Larson, 1971; Maltz and McCleary, 1977). When measuring the recidivism of each defendant, we included only arrests that occurred after the disposition date in the sampled case—the date of conviction, acquittal, or case dismissal (not the date of sentencing). Our measure of recidivism was a binary variable that took a value of one if the defendant was rearrested on any criminal charge (felony or misdemeanor, drug-related or not, including domestic violence charges) in the District of Columbia or Maryland within 4 years after the disposition of the sampled case. In other words, regardless of whether defendants were released or imprisoned, we tracked them for 4 years after their disposition dates. As noted in the subsequent discussion, we also performed robustness checks, using alternative definitions of what types of outcomes count as recidivism.

It might seem odd to start the clock at the point of disposition rather than at release, but this approach preserves the symmetry between defendants randomly assigned to different judges. These groups are identical in expectation, the sole difference being the sentence that was meted out. To start the clock at the time of release would confound the effect of incarceration with the effect of age (as well as any other time-related factor) because defendants assigned to harsh judges begin their terms of release at an older age, on average, than their control group counterparts. This confound would undercut our ability to draw causal inferences about the effects of punishment. Starting the clock at the point of case disposition preserves our ability to draw causal inferences about the overall effect of punishment; the problem is that the specific deterrence mechanism is intertwined with the incapacitation mechanism. In our application, this problem was mitigated by the fact that almost every defendant had ample time to recidivate after release. In the end, we found weak effects of punishment on recidivism, which implies that specific deterrence has little influence on criminal behavior.

COVARIATES

Although not necessary for unbiased inference, the covariates available in public records serve two statistical purposes. First, the background attributes of defendants help verify the random assignment of judges. As we later demonstrate, we found no systematic relationship between the criminal backgrounds of defendants and the judicial calendar to which they were assigned. Second, covariates may reduce unexplained variability in recidivism, thereby improving the precision with which we estimated the effects of our sentencing variables.

Our coding of the covariates was as follows. We employed both a linear and a squared term for age, which we computed as the year of the defendant's arrest minus the defendant's year of birth. We used a binary categorical variable for race, which takes a value of one if the defendant is not Black. District of Columbia court records seem not to distinguish between Hispanics and Caucasians. For gender, we used a binary categorical variable that took a value of one for female defendants. To measure the criminal history of a defendant, we used an extensive battery of dummy variables, each marking one of the following attributes: prior arrest, prior drug arrest, prior felony arrest, prior felony drug arrest, prior conviction, prior drug conviction, prior felony conviction, and prior felony drug conviction. Thus, a defendant who had a prior felony drug arrest and conviction would be scored one on each of these dummy indicators. Finally, we included a marker for the charged offense. We used two dummy variables for a PWID charge and a distribution charge. All defendants in the sample had at least one of these charges; some had both. We used six binary categorical variables to classify the drug or drugs that the defendant allegedly possessed or distributed—marijuana, cocaine, crack cocaine, heroin, PCP, or "other" (prescription drugs, ecstasy, etc.). Finally, a binary variable denotes whether the defendant faced any nondrug charges (of the nature discussed earlier).[8] It should be stressed that, in contrast to models of observational data, it is not imperative to have the "correct" set of covariates to obtain consistent estimates from the instrumental variables estimator. Our list of covariates doubtless excludes a range of personal and contextual factors that could affect recidivism, but these factors are statistically independent of judge assignment. Similarly, our covariates include several overlapping measures of past criminal activity, but our aim was not to estimate the unique impact of each; instead, we sought to control for an array of factors that *jointly* predict recidivism.

Because codefendants are assigned to the same judge, their observations cannot be considered independent for the purpose of calculating standard errors (Arceneaux, 2005). We identified 172 codefendants by using the complaint control number (CCN) recorded on the lockup list. The CCN denotes a particular instance of criminal activity and, therefore, can be associated with multiple individual arrests; codefendant arrests have the same CCN. The regression analyses discussed in the following section use robust cluster standard errors to account for the nonindependence of codefendant observations.

JUDGE ASSIGNMENTS

Felony cases entering the Court are placed on one of the following dockets: Felony I, Felony II, or the Accelerated Felony Trial Calendar (AFTC). The Felony I docket hears cases of murder and sexual assault. The AFTC docket hears cases, generally crimes of violence, in which the defendant has been detained preventively and must be tried within 100 days. The Felony II docket hears all the rest, including most drug felony cases (USGAO, 2001). Drug cases occasionally do appear on the AFTC docket if the defendant has been detained preventively, but such appearances are rare and we excluded them from our analysis. The exclusion does not introduce bias because the assignment of a case to the AFTC docket precluded its assignment to a Felony II calendar.

We focused on cases placed on the Felony II docket. Once placed on this docket, cases are assigned to one of several calendars that serve the docket. During 2002 and 2003, the period from which our sample was drawn, the Court used a mechanical wheel to rotate the assignment of new cases among the calendars—assigning one case to calendar 1, the next case to calendar 2, and so on. The arraignment court coordinator in March 2007 explained that she deviated from the cycle when the caseload of a calendar

8. We used arrest charges rather than charges at disposition or sentencing because only arrest charges are independent of judge assignment. The charges at sentencing might reflect plea bargains, which are shaped heavily by judges' reputations (Lacasse and Payne, 1999; see also Landes, 1971).

was out of balance with the rest, generally because the judge in question had processed cases faster or slower than the norm. When such imbalances occur, she explained, the coordinator can skip an overloaded calendar in the cycle or assign additional cases to an underloaded one. It also seems that defendants who have cases pending before a particular judge have additional cases assigned to the same judge. The process remains random insofar as the assignments never depend on the facts of the case or properties of the defendant.[9] Cases remain on the same calendar through the final disposition, but the judges assigned to each calendar sometimes rotate at the beginning of each year. We, therefore, considered calendar assignment rather than specific judge assignment to be the randomly assigned treatment. In effect, the random assignment of calendars causes defendants to be exposed to different sets of judges.

Because the court does not use a random number generator and some degree of discretion is placed in the hands of the clerk in charge of calendar assignments, it is important to verify that the assignment process that was implemented had statistical properties consistent with a process of true random assignment. Table 2 compares selected covariates across calendars and shows them to be well balanced. To assess the distribution of cases statistically, we conducted a multinomial logistic regression analysis of calendar assignment on each of the covariates listed in table 2. Under the null hypothesis, these predictors bore no systematic relationship to calendar assignment. This hypothesis was evaluated by means of a chi-square test, the size of which was determined by Monte Carlo simulations.[10] As expected, this test proved nonsignificant for each of the covariates and for all of the covariates considered jointly (p = .96). Because the covariates were balanced across judicial calendars, the inclusion of covariates had little effect on the regression estimates reported in the following discussion.

Table 2 also conveys important information about the kinds of defendants in our sample. Particularly noteworthy is the fact that 85 percent of the defendants in our sample had at least one prior arrest and that 67 percent had at least one prior conviction. In other words, we are studying the deterrent effects of punishment on a set of individuals for whom past

9. We verified this information by telephone with Alicia Shepard, Arraignment Court Coordinator, as of March 2007.

10. The nominal chi-square thresholds become severely biased in favor of rejection as the number of degrees of freedom increases (Hansen and Bowers, 2008). Monte Carlo simulations randomly assigned observations (in clusters, where codefendants were concerned) to calendars and computed the chi-square from each multinomial logit regression. This procedure was repeated 1,000 times to obtain the distribution of the chi-square statistics under the null hypothesis.

Table 2. Defendant Characteristics, by Calendar Assignment

	Calendar 1	2	3	4	5	6	7	8	9	*p* Value
Age	31.9	35.1	33.2	32.8	33.8	32.2	33.3	34.2	32.3	.62
	(11.5)	(11.8)	(11.6)	(10.8)	(11.1)	(11.1)	(11.3)	(11.5)	(10.6)	
Female	13.1	7.1	7.6	10.5	9.5	8.6	10.1	9.1	11.8	.93
Non-Black	4.1	4.5	1.7	3.2	.9	2.2	1.8	1.0	2.7	.84
Prior arrest	81.1	86.6	85.6	83.1	87.1	81.7	78.9	90.9	93.6	.07
Prior drug arrest	68.0	74.1	74.6	71.8	80.2	64.5	66.1	73.7	75.5	.34
Prior felony arrest	63.1	73.2	70.3	74.2	75.9	67.7	70.6	72.7	79.1	.41
Prior felony drug arrest	54.1	58.9	59.3	57.3	59.5	48.4	45.9	56.6	56.4	.52
Prior conviction	59.8	69.6	64.6	71.0	72.4	67.7	62.4	66.7	70.9	.54
Prior drug conviction	50.0	53.6	52.5	58.1	66.4	54.8	47.7	53.5	57.3	.30
Prior felony conviction	43.4	58.9	55.1	54.0	59.5	50.5	50.5	54.6	59.1	.34
Prior felony drug conviction	35.3	44.6	47.5	44.4	50.0	39.8	34.9	44.4	43.6	.36
PWID charge	49.2	40.2	49.2	41.1	56.0	50.5	52.3	43.4	39.1	.25
Distribution charge	59.8	68.8	61.9	68.6	52.6	59.1	54.1	61.6	67.3	.20
Marijuana charge	22.1	17.0	17.0	17.0	23.3	18.3	17.4	18.2	20.9	.95
Cocaine charge	39.3	38.4	45.8	40.3	33.6	40.9	44.0	33.3	43.6	.75
Crack cocaine charge	14.8	15.2	18.6	19.4	20.7	23.7	24.8	22.2	19.1	.75
Heroin charge	23.8	31.3	29.7	25.8	30.2	29.0	15.6	29.3	22.7	.34
PCP charge	6.6	7.1	4.2	2.4	6.0	1.1	6.4	4.0	3.6	.52
Other drug charge	4.9	.0	3.4	4.0	3.5	2.2	4.6	3.0	3.6	.55
Nondrug charge	11.5	8.9	17.0	12.9	12.9	10.8	15.6	12.1	14.6	.81
n	122	112	118	124	116	93	109	99	110	

NOTES: Total N = 1,003. Entries are means (age) and percentages. For continuous variables, standard deviations are in parentheses. The p values in the final column refer to the significance of a multinomial regression in which judge calendar assignment was regressed on each variable individually. These p values were obtained from Monte Carlo simulations as explained in the text.

interactions with the criminal justice system largely have failed to deter subsequent criminality.

STATISTICAL MODEL

In recent years, several scholars have used random judge assignments to assess the downstream consequences of sentencing on employment and earnings (Kling, 2006; Waldfogel, 1994) as well as recidivism (Berube and Green, 2007). The logic of our study follows the same framework as that of the Waldfogel (1994), Kling (2006), and Berube and Green (2007) studies, which in turn, drew on the statistical results presented in Angrist, Imbens, and Rubin (1996) and in Imbens and Angrist (1994). Rather than recapitulate these models, we summarize the key theorem and the assumptions on which it is based. We then discuss the empirical adequacy of these assumptions given the data at hand.

Suppose, for simplicity, that we have two judges, denoted $Z = \{0,1\}$. A defendant is assigned at random to one of them. Again, for simplicity, assume that a judge makes a binary decision whether to incarcerate. This decision X takes the value 1 for incarceration and 0 otherwise. A period of time elapses since this decision, and we observe whether the defendant has been rearrested.[11] Let Y take the value of 1 if a rearrest occurred and 0 otherwise. The causal effect of sentencing on recidivism is, in principle, the difference between two states of the world, one in which the defendant was incarcerated and one in which the defendant was released. Because we only observe one such outcome for each defendant, we must devise a way to draw inferences about the average causal effect in the population of defendants. Angrist, Imbens, and Rubin (1996) showed that when certain assumptions are met, random assignment creates comparable groups whose outcomes may be used to estimate an average causal effect.

Let us specify more precisely what these assumptions are and what kind of causal effect is estimated. First, Z must exert a nonzero effect on X, which is to say that judges must have different sentencing propensities. This assumption may be assessed empirically by examining whether judges' sentences vary more than would be expected by chance. Of course, in any given sample, one set of judges might be randomly assigned more serious crimes or more pathological defendants; the question is whether variation in sentencing exceeds what one would expect given random assignment. Second, one must assume that the potential outcomes for the defendant are independent of judge assignment, which is satisfied given

11. Some studies of recidivism have compared rearrest rates for a certain period of time after release from custody, but this practice introduces the risk of bias because the defendants are no longer equivalent in terms of age and are no longer presented with the same environment within which to recidivate.

the random assignment of judges. Third, one must assume that a defendant's judge influences recidivism only through the sentence he or she hands down. Any other influence of a particular judge—for instance, the statements that she makes at sentencing—must be inconsequential. Fourth, one must assume that sentences handed down to one defendant have no direct causal influence on outcomes associated with another defendant. This so-called stable unit treatment value assumption supposes that it is inconsequential whether defendants compare their sentences; in the end, they are affected only by their own sentences. The stable unit treatment value assumption also raises the question of whether treatment effects are constant across defendants. If not, the fifth assumption becomes operative. It states that judges' punitiveness may be rank ordered and that this rank ordering is preserved across all defendants they confront. Under these assumptions, Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996) showed that an instrumental variables regression of Y on X, using Z as an instrument for X, provides consistent estimates of the local average treatment effect (LATE). The LATE in our example is the average causal effect of incarceration on recidivism among those who would be incarcerated by the more punitive judge but not by the less punitive judge.

The first assumption was clearly met in our data. We had nine court calendars to which cases were assigned randomly, and sentences varied substantially across calendars. As shown in table 3, the least punitive calendar put 23 percent of defendants behind bars, whereas the most punitive calendar incarcerated 65 percent of defendants. A wide variation also was observed in the use of probation. One calendar assigned probationary periods to 29 percent of its defendants, whereas another gave 60 percent of its defendants probation. Fortunately, for the purpose of inferring the distinct causal role of incarceration and probation, the two variables were modestly correlated (–.08) when the data were aggregated at the calendar level.

Variation in sentencing across calendar assignments far exceeded what could be expected by chance. As shown in table 4, when we regressed whether a defendant was incarcerated on dummy variables marking eight of the nine calendars, we found the joint significance of these regressors to be $p < .001$ based on an F test. The same was true when we regressed prison sentence length on the eight calendar dummies. Again, an F test showed the calendar assignments to be significant predictors of sentence length, with a p value of less than .01. Calendar assignments also significantly predicted whether a defendant was sentenced to probation ($p < .001$) as well as the length of the probation period ($p < .001$). Although

Table 3. Prison Sentences, Probation Sentences, and Recidivism Rates, by Calendar

	Calendar								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
% incarcerated	33.6	25.0	65.4	56.5	33.6	48.4	43.1	23.2	44.5
Average nonsuspended prison sentence (in months)	5.1	7.6	11.9	7.8	5.8	5.1	5.6	5.5	7.1
% sentenced to probation (instead of or in addition to prison)	50.0	57.1	42.4	31.5	43.1	60.2	29.4	48.5	32.7
Average probation sentence (in months)	12.5	11.5	11.7	6.8	13.7	14.9	6.4	8.7	7.1
% arrested within 4 years of case disposition	48.4	45.5	57.6	49.2	56.0	58.1	49.5	44.4	65.5
n	122	112	118	124	116	93	109	99	110

studies of randomly assigned judges may founder if the sentencing patterns of judges differ to a minor degree, our data showed substantial variation across calendars—a fact that contributes to the precision with which we later estimate the effects of sentences on recidivism.

The second assumption is equivalent to an "exclusion restriction" in a simultaneous equation model. As such, it cannot be tested directly. In this application, however, it seems intuitive that sentence severity is the only way that judge assignments could affect recidivism. The question is how to measure the severity of the sentence. Some scholars look at whether defendants were incarcerated, whereas others look at the duration of prison time. Probation time is properly considered part of the sentence, and one could code probation in terms of whether it is assigned or the length of the probation period. Our approach was to focus primarily on the length of incarceration and probation, but we also assessed the robustness of our findings using alternative conceptualizations of punishment and obtained similar results.

The stable unit treatment value assumption cannot be assessed directly, but again, it seems plausible in this application. If defendants are subject to what has been termed "specific deterrence," then it is presumably because their personal experience with punishment sensitizes them to the downside risks of future criminal conduct. Granted, the District of Columbia is a relatively small jurisdiction, and it is possible that defendants know each other and perhaps even know each other's sentences. This information does not necessarily impair the internal validity of the study, although

Table 4. *F* Tests of the Joint Significance of Calendar Assignment on Incarceration and Probation Sentences

Dependent Variable	Covariates	Numerator Degrees of Freedom	Denominator Degrees of Freedom	*F*
Incarceration	No	8	994	9.80**
Incarceration	Yes	8	973	10.99**
Prison term	No	8	994	3.19*
Prison term	Yes	8	973	3.09*
Probation	No	8	994	5.65**
Probation	Yes	8	973	5.92**
Probation term	No	8	994	6.63**
Probation term	Yes	8	973	7.10**

NOTES: Incarceration and probation are binary variables scored 1 if the defendant were imprisoned or sentenced to probation, respectively. Prison term and probation term were coded in terms of month of sentence. Prison term refers only to nonsuspended prison time. Covariates include age, age-squared, sex, race, prior arrest, prior drug arrest, prior felony arrest, prior felony drug arrest, prior conviction, prior drug conviction, prior felony conviction, prior felony drug conviction, charge of possession with intent to distribute, and charge of distribution, and dummy variables for possession of marijuana, cocaine, crack cocaine, heroin, or PCP; other drug charges; and nondrug charges. * $p < .01$, ** $p < .001$, with or without correction for clustering.

it might circumscribe its external validity. Suppose that defendants were emboldened to recidivate upon learning that other defendants received light sentences. This knowledge in itself would not bias our results unless defendants assigned to different judges were exposed differentially to this information. If defendants have an equal probability of encountering this criminogenic bit of news, then it effectively becomes part of the overall context within which the experiment takes place. Information diffusion might explain why specific deterrence fails in a setting where light sentences are handed out, but it does not cause us to misestimate the effects of sentence severity in such a context.

Finally, we come to the question of whether the severity of judges can be ranked in a monotonic fashion. This assumption again cannot be assessed directly, but certain features of our design bolster its plausibility. Our sample is drawn from a narrow class of criminal cases all involving drugs. None of the cases involves other kinds of felonies. Thus, we need not worry that some judges take a stern view of property crime or violent crime but look the other way when it comes to drug-related crime. Because all of our cases fall within a narrow range of criminal activity, it is harder to imagine idiosyncratic sentencing philosophies that could lead to violations of monotonicity. Furthermore, the fact that all of our cases were heard within a short span of time makes it unlikely that the sentencing philosophies of judges changed appreciably during that period.

For these reasons, it is plausible to think that the instrumental variables estimator will generate meaningful estimates of what Angrist, Imbens, and Rubin (1996) term the local-average treatment effect of sentencing on recidivism. As shown in table 3, drug defendants assigned to the most punitive calendar of judges served an average (nonsuspended) prison sentence of 11.9 months, whereas defendants assigned to the least punitive calendar of judges served 5.1 months. The least punitive group of judges sentenced 23 percent of its defendants to prison as opposed to 65 percent of those whose cases were heard by the most punitive calendar of judges. In essence, our experiment allows us to assess the average causal effects of randomly doubling prison terms or randomly deciding whether those falling between the 23rd and 65th percentiles of culpability go to prison.

The estimator itself is simply a multivariate generalization of the instrumental variables regression model presented in Angrist, Imbens, and Rubin (1996). Recidivism (Y_i) is modeled as a linear function of the two endogenous treatments, incarceration (I_i) and probation (P_i), each expressed in terms of months sentenced. This model may be augmented with the set of covariates mentioned earlier—demographic variables (age, age-squared, sex, and race), prior record variables (arrest, drug arrest, felony arrest, felony drug arrest, conviction, drug conviction, felony conviction, and felony drug conviction), charge variables (possession with intent to distribute and distribution, and nondrug charges), and drug type (marijuana, cocaine, crack cocaine, heroin, PCP, and other drugs). For each dummy variable, defendants were scored 1 if they fell into the applicable category and 0 otherwise. It should be stressed that each of these covariates uses information observed prior to judge assignment and that the extensive set of covariates related to past criminal activity compares favorably with other studies of recidivism among drug offenders (cf. Kim et al., 1993).

Denoting these covariates as the matrix X, whose effects are the vector Γ, we may write the model as follows:

$$Y_i = \beta_0 + \beta_1 I_i + \beta_2 P_i + X\Gamma + u_i \quad (1)$$

where u_i represents unobserved disturbances affecting recidivism. The key parameters are β_1 and β_2, the effects of incarceration and probation, respectively. Because the disturbances likely will be correlated with incarceration and probation, we estimated the parameters of this model using two-stage least-squares regression in which calendar assignment provided the excluded instrumental variables. Specifically, we used as instrumental variables eight dummy variables $\{C_{1i}, C_{2i}, ..., C_{8i}\}$, one marking each of the calendars, less one. With two endogenous regressors and eight excluded instruments, this model was overidentified.

RESULTS

As a preliminary step, we estimated a reduced-form regression as follows:

$$Y_i = \gamma_0 + \gamma_1 C_{1i} + \gamma_2 C_{2i} + \ldots + \gamma_8 C_{8i} + X\Pi + \varepsilon_i \qquad (2)$$

This exercise provides a straightforward indication of whether random calendar assignments have a downstream effect on recidivism. The advantages of this approach are that it makes minimal assumptions about the causal paths through which judicial sentences influence recidivism and is unaffected by the issue of weak instruments (Chernozhukov and Hansen, 2008), which we discuss in the next section. The results suggest that recidivism rates vary across calendars, although the relationship is of borderline statistical significance. An *F* test of the joint significance of the estimates of $\{\gamma_1, \gamma_2, ..., \gamma_8\}$ has a *p* value of .025 without covariates and of .040 controlling for covariates. These preliminary results suggest that calendar assignments might exert some causal influence on recidivism, but the influence is fairly weak, which implies that deterrence and incapacitation play fairly weak mediating roles.

Table 5 presents six different variants of equation 1. The variants allowed us to examine the sensitivity of the results to the inclusion of covariates.[12] We also examined whether the results change appreciably according to whether incarceration and probation are included together in the same regression. Two conclusions emerge from this array of estimates. First, incarceration seems to have weak effects on recidivism. The strongest effect we estimated for months of incarceration was .009 (see columns 2 and 6 of table 5), which implies, *ceteris paribus,* that a 16.5-month prison sentence (the average sentence among those who are imprisoned) produces a (16.5)x(.009)=14.9-percentage-point gain in the expected probability of recidivism. The weakest estimate, .008 (see columns 1 and 5 of table 5), implies that the average sentence increases the probability of recidivism by 13.2 percentage points. We regard these effects as substantively small. In none of the specifications does the estimated effect of imprisonment approach conventional levels of statistical significance.

12. Because the covariates were not assigned randomly, the coefficients associated with the covariates have no direct causal interpretation. Prior felony drug arrests, for example, might cause recidivism or might be markers for unobserved attributes that predict recidivism. Moreover, our covariates were designed to be coded as flexibly as possible, and the inclusion of both prior arrests and prior convictions means that these variables are redundant indicators of unobserved propensity to reoffend. This issue undercuts our ability to interpret the covariates in causal terms but does not impair our ability to draw causal inferences about randomly assigned sentences, which is our principal aim.

Table 5. 2SLS Estimates of the Effects of Length of Incarceration and Probation on Recidivism

Independent Variables	Models (1)	(2)	(3)	(4)	(5)	(6)
Incarceration (in months)	.008 [.008]	.009 [.008]			.008 [.008]	.009 [.008]
Probation (in months)			.002 [.005]	.001 [.005]	.003 [.006]	.002 [.005]
Age		–.025* [.010]		–.025* [.010]		–.024* [.010]
Age-squared		.000 [.000]		.000 [.000]		.000 [.000]
Female		–.001 [.064]		–.032 [.058]		–.004 [.064]
Non-Black		–.219 [.112]		–.191 [.102]		–.211 [.117]
Prior arrest		–.060 [.073]		–.070 [.072]		–.061 [.074]
Prior drug arrest		.007 [.065]		.003 [.063]		.008 [.065]
Prior felony arrest		.104 [.066]		.137* [.065]		.112 [.070]
Prior felony drug arrest		–.100 [.068]		–.119* [.066]		–.103 [.068]
Prior conviction		.020 [.072]		.031 [.069]		.021 [.072]
Prior drug conviction		.041 [.073]		.056 [.069]		.039 [.073]
Prior felony conviction		–.100 [.072]		–.083 [.068]		–.105 [.074]
Prior felony drug conviction		.056 [.082]		.088 [.079]		.063 [.085]
PWID charge		.011 [.063]		.015 [.060]		.008 [.063]
Distribution charge		.011 [.066]		.023 [.068]		–.001 [.076]
Marijuana		.100 [.055]		.087 [.057]		.093 [.060]
Cocaine		–.000 [.058]		–.004 [.060]		–.008 [.062]
Crack		.040 [.064]		.031 [.066]		.031 [.069]
Heroin		.084 [.061]		.075 [.061]		.077 [.065]
PCP		.082 [.094]		.111 [.085]		.077 [.095]
Other drug		–.040 [.109]		–.058 [.104]		–.046 [.112]
Nondrug charge		.001 [.048]		.011 [.045]		.003 [.048]
Constant	.471* [.058]	1.012* [.183]	.504* [.058]	1.015* [.191]	.443* [.084]	.986* [.202]
p value associated with *F* test of joint significance of all regressors	.324	<.001	.687	<.001	.551	<.001

NOTES: N = 1,003. All specifications use calendar dummy variables as excluded instrumental variables. Numbers in brackets are robust cluster standard errors, which take into account a few cases in which codefendants were assigned to the same calendar. Recidivism is defined as rearrest within 4 years after a defendant's case disposition.
* p < .05 (two-tailed).

Second, probation seems to do little to reduce the probability of recidivism. The estimates reported in table 5 are only mildly positive. The strongest coefficient, .003 (column 5 of table 5) implies that an average probation sentence among those who receive probation (23.7 months) produces a 7.1 percentage-point increase in the rate of recidivism. It should be stressed, however, that neither this estimate nor the other estimated effects of probation reported in table 5 are remotely close to statistical significance. Overall, the instrumental variables regression results, like the reduced-form regression results, suggest that sentences weakly influence recidivism. *Ceteris paribus*, the median defendant who experiences both incarceration and probation is expected to recidivate at approximately the same rate as a defendant released without punishment or supervision.

ASSESSING ROBUSTNESS

Instrumental variables estimators are designed to counteract the biases associated with endogenous treatment—in this case, the fact that the terms of incarceration and probation might reflect defendant attributes that predict recidivism. One question is how the results would look if instead we were to use ordinary least-squares (OLS) regression on the assumption that sentences are meted out in a fashion that is random conditional on the covariates listed in table 2. This assumption is dubious because judges have access to information about defendants that goes beyond the list of covariates in our model (Achen, 1986) and because defendants might plea bargain strategically in light of the judge they receive (Lacasse and Payne, 1999).

Table 6 reports the results of a series of OLS regressions that parallel the two-stage least-squares (2SLS) regressions reported in table 5. Although the 2SLS estimates of incarceration are positive and insignificant, the OLS estimates are negative and highly significant. The second column of table 6, for example, suggests that each month of incarceration lowers the probability of recidivism by .006, with a standard error of just .001. The OLS results for probation, like the 2SLS results, are weak and insignificant. Overall, however, the implication of a conventional OLS regression analysis is that prison time significantly reduces the probability of reoffending.

The fact that the IV and OLS estimators produce different estimates is subject to two competing interpretations. The first is that OLS is biased because judges are assigned at random but sentences are not. In particular, longer prison terms might be meted out to defendants who happen to be reaching the end of lengthy criminal careers. This argument receives some support from a Hausman test comparing the 2SLS results with the OLS results; both estimators are consistent under the null hypothesis that

Table 6. OLS Estimates of the Effects of Length of Incarceration and Probation on Recidivism

Independent Variables	Models					
	(1)	(2)	(3)	(4)	(5)	(6)
Incarceration (in months)	–.005*	–.006*			–.005*	–.005*
	[.001]	[.001]			[.001]	[.001]
Probation (in months)			.002	.001	.001	.000
			[.001]	[.001]	[.001]	[.001]
Age		–.027*		–.025*		–.026*
		[.009]		[.009]		[.009]
Age-squared		.000		.000		.000
		[.000]		[.000]		[.000]
Female		–.047		–.032		–.048
		[.057]		[.058]		[.058]
Non-Black		–.184		–.190		–.182
		[.095]		[.100]		[.095]
Prior arrest		–.075		–.070		–.075
		[.072]		[.073]		[.072]
Prior drug arrest		–.000		.003		–.000
		[.064]		[.064]		[.064]
Prior felony arrest		.148*		.138*		.150*
		[.062]		[.062]		[.062]
Prior felony drug arrest		–.127		–.120		–.128
		[.066]		[.066]		[.066]
Prior conviction		.037		.031		.037
		[.070]		[.070]		[.070]
Prior drug conviction		.068		.056		.067
		[.069]		[.070]		[.069]
Prior felony conviction		–.066		–.083		–.067
		[.067]		[.067]		[.067]
Prior felony drug conviction		.099		.088		.100
		[.076]		[.077]		[.076]
PWID charge		.020		.014		.020
		[.059]		[.060]		[.059]
Distribution charge		.043		.021		.041
		[.059]		[.061]		[.060]
Marijuana		.087		.086		.086
		[.052]		[.053]		[.053]
Cocaine		.003		–.004		.002
		[.056]		[.056]		[.056]
Crack		.037		.030		.035
		[.061]		[.062]		[.062]
Heroin		.078		.075		.077
		[.058]		[.059]		[.058]
PCP		.135		.111		.134
		[.085]		[.085]		[.085]
Other drug		–.061		–.058		–.062
		[.101]		[.104]		[.101]
Nondrug charge		.016		.012		.016
		[.045]		[.046]		[.045]
Constant	.560*	1.047*	.507*	1.012*	.550*	1.042*
	[.018]	[.172]	[.020]	[.175]	[.023]	[.174]
R^2	.017	.091	.003	.073	.017	.091

NOTES: N = 1,003. All specifications use calendar dummy variables as excluded instrumental variables. Numbers in brackets are robust cluster standard errors, which take into account a few cases in which codefendants were assigned to the same calendar. Recidivism is defined as rearrest within 4 years after a defendant's case disposition.
* $p < .05$ (two-tailed test).

sentences are unrelated to omitted determinants of recidivism, but OLS is inconsistent under the alternative hypothesis (Hausman, 1978). Comparing the estimated effects of incarceration in column 2 of tables 5 and 6, the Hausman test nears significance (x^2 with 1 d.f. = 3.05, $p < .1$). Tests involving probation or probation and incarceration are nonsignificant ($p > .25$) because probation has negligible effects regardless of the estimation technique.

An alternative interpretation is that the instrumental variables estimates are biased. Even when instruments are valid (i.e., independent of unobserved factors affecting recidivism), they might be weak in the sense that they predict relatively little variance in the endogenous explanatory variables. Bound, Jaeger, and Baker (1995) pointed out that weak instruments may lead to inconsistent parameter estimates that are biased in the direction of the OLS estimates. Staiger and Stock (1997) suggested that first-stage *F* statistics for excluded instrumental variables should be greater than ten—a result generalized by Stock and Yogo (2002). The *F* statistics in table 4 fall short of this criterion, although it should be noted that this rule of thumb is conservative in that it is designed to rule out the possibility that weak instrument bias exceeds 10 percent of the bias associated with OLS (Stock and Yogo, 2002: 32). Following the advice of Stock and Watson (2007: 441) and Angrist and Pischke (2009: 213), we experimented with 2SLS specifications that exclude all but the strongest instruments (i.e., dummy variables marking the judges with the most distinctive sentencing patterns) and obtained first-stage *F* statistics in excess of ten but obtained almost identical 2SLS estimates of the effects of incarceration and probation.[13]

The econometric literature on weak instruments suggests that limited information maximum likelihood (LIML) estimation "has the advantage of having the same asymptotic distribution as 2SLS . . . while providing a finite-sample bias reduction" (Angrist and Pischke, 2009: 209). Table 7 presents LIML estimates using the same model specification as in table 5. The LIML estimates parallel the 2SLS results, with somewhat larger positive estimates for incarceration and larger standard errors. For example, column 2 of table 7 suggests that each month of incarceration increases the probability of recidivism by .021, with a standard error of .018.

13. These supplementary results are presented in an online appendix at http://vote.research.yale.edu/replication.html.

Table 7. Limited Information Maximum Likelihood Estimates of the Effects of Length of Incarceration and Probation on Recidivism

Independent Variables	Models					
	(1)	(2)	(3)	(4)	(5)	(6)
Incarceration (in months)	.022 [.020]	.021 [.018]			.021 [.019]	.021 [.018]
Probation (in months)			.002 [.008]	.001 [.007]	.002 [.008]	.001 [.007]
Age		–.025* [.011]		–.025* [.010]		–.024 [.012]
Age-squared		.000 [.000]		.000 [.000]		.000 [.000]
Female		.039 [.083]		–.032 [.059]		.036 [.084]
Non-Black		–.250 [.144]		–.191 [.105]		–.242 [.152]
Prior arrest		–.047 [.078]		–.070 [.072]		.048 [.078]
Prior drug arrest		.014 [.071]		.003 [.063]		.014 [.071]
Prior felony arrest		.067 [.082]		.137 [.069]		.073 [.089]
Prior felony drug arrest		–.077 [.077]		–.119 [.066]		–.080 [.078]
Prior conviction		.006 [.078]		.031 [.069]		.007 [.078]
Prior drug conviction		.017 [.087]		.056 [.069]		.016 [.086]
Prior felony conviction		–.130 [.090]		–.082 [.070]		–.133 [.092]
Prior felony drug conviction		.018 [.100]		.087 [.082]		.025 [.105]
PWID charge		.004 [.071]		.015 [.060]		.001 [.072]
Distribution charge		–.017 [.084]		.023 [.075]		–.026 [.097]
Marijuana		.111 [.062]		.087 [.061]		.105 [.070]
Cocaine		–.003 [.065]		–.003 [.063]		–.010 [.072]
Crack		.043 [.071]		.032 [.071]		.035 [.080]
Heroin		.088 [.070]		.076 [.064]		.082 [.076]
PCP		.036 [.118]		.111 [.086]		.033 [.119]
Other drug		–.022 [.129]		–.057 [.106]		–.028 [.132]
Nondrug charge		–.011 [.054]		.011 [.046]		–.009 [.055]
Constant	.377* [.137]	.982* [.211]	.503* [.083]	1.016* [.205]	.357* [.153]	.960* [.241]
p value associated with *F* test of joint significance of all regressors	.273	<.001	.770	<.001	.518	<.001

NOTES: N = 1,003. Following the same procedures as in table 5 but using LIML instead of 2SLS, all specifications use calendar dummy variables as excluded instrumental variables. Numbers in brackets are robust cluster standard errors, which take into account a few cases in which codefendants were assigned to the same calendar. Recidivism is defined as rearrest within 4 years after a defendant's case disposition. * p < .05 (two-tailed test).

Finally, we investigated an array of different modeling and measurement strategies. We considered, for example, restricting the definition of recidivism to subsequent drug arrests, felony arrests, and nonfelony arrests. Regardless of the way in which recidivism was measured, the weak and insignificant results reported in table 5 hold up. We also looked for interactions between sentencing and criminal history. For example, we partitioned the data set according to whether defendants had a prior conviction on the grounds that those confronting criminal penalties for the first time might be more susceptible to deterrence (Kim et al., 1993; Witte, 1980). Again, we did not find evidence of differential effects when partitioning the data in this manner. In addition, we looked for possible interactions between the length of the suspended incarceration time imposed on a defendant and the length of the probation sentence for that defendant. Because those who violate probation may incur the incarceration time that previously had been suspended, a lengthier period of suspended incarceration increases the expected costs of probation violation. No such interaction was found, however. Also, at the suggestion of a reviewer, we excluded defendants who were not convicted. Ordinarily, one would not restrict the sample based on a variable that is arguably the consequence of judicial assignment. Differential conviction rates could reflect differential sentencing tendencies among the judges, in the sense that defendants facing harsh expected penalties for going to trial might be more likely to plead guilty (see Lacasse and Payne, 1999), and controlling for conviction might introduce what Gelman and Hill (2007: chapter 9) call posttreatment bias. It turns out that the results excluding the nonconvicted defendants continue to show weak and nonsignificant effects for incarceration and probation. These supplementary results are included in the online appendix.

DISTINGUISHING SPECIFIC DETERRENCE FROM INCAPACITATION

As noted, our research design allows us to estimate the combined effects of specific deterrence and incapacitation, not the distinct effects of each. For policy purposes, the combined effect might be of primary interest; for scholarly purposes, the distinct effect of specific deterrence is often what researchers seek to estimate (e.g., Martin, Annan, and Forst, 1993). Although the role of specific deterrence cannot be assessed without invoking additional assumptions, it is possible to obtain an approximate sense of its influence by way of a statistical simulation.

The purpose of our simulation is to isolate the effect of specific deterrence by eliminating the effects of incapacitation. In other words, we simulate how incarcerated defendants would have behaved if they were not behind bars. To do so, we started with the release date of each defendant, assuming that he serves 100 percent of his nonsuspended prison sentence beginning on the date of disposition.[14] We estimated the hazard rate of recidivism as a function of time since this putative release date. The survival model is a Weibull regression in which the predictors are the covariates listed in tables 5–7. Using the estimated hazard rate parameters, we then calculated the predicted survival probability of each defendant, x, at the end of a full 4-year recidivism window. Let $p = 1 - x$ be the probability that a given defendant recidivates at least once during the 4-year window. Our simulation substituted a Bernoulli random variable that takes value 1 with probability p in place of observed recidivism for defendants who were incarcerated and did not recidivate during their shortened period of time on the street. Modeling incapacitation in this fashion raises the rate of recidivism from 53 percent (observed) to 62 percent (simulated). This approach, of course, assumes that the hazard function for those out of prison is the same as the hazard function for the incarcerated had they been out of prison.

Reestimating the 2SLS regression in column 2 of table 5 with simulated recidivism as the dependent variable, we found the average effect of incarceration (in months) to be .0170 with a standard error of .0082 across 1,000 simulations. The corresponding LIML estimate is .0302 with a standard error of .0186.[15] Both the 2SLS and LIML results imply that, net of incapacitation, longer sentences lead to higher rates of recidivism among drug offenders. The statistical significance of these positive estimates is marginal; approximately 84 percent of the simulated samples show a significantly positive ($t > 1.65$) effect using 2SLS, whereas 51 percent show a significant positive effect using LIML. In sum, after bracketing the effects of incapacitation, the statistical results provide some suggestive evidence for the hypothesis that incarceration has criminogenic consequences. Although the criminogenic effect remains somewhat speculative, these simulation results certainly cast doubt on the hypothesis that punishment exerts a specific deterrent effect.

CONCLUSION

The random assignment of defendants to judges of varying punitive styles sets the stage for a natural experiment with important implications for both policy and behavioral science. The subsequent criminal histories

14. This coding decision reflects our assumption that many defendants who incur prison time are held in custody from the time of disposition, prior to their sentencing.
15. These estimates are reported in the online appendix.

of more than 1,000 defendants arrested on drug-related charges in the District of Columbia support two provocative findings. The first is that incarceration seems to have little net effect on the likelihood of subsequent rearrest. Despite the fact that we measured recidivism in a way that gives those incapacitated by prison less time to recidivate than those who are not incarcerated, prison time seems to do little to reduce the odds of rear-rest. Evidently, the combined effects of incapacitation and specific deterrence are weak in this setting.

Perhaps this deterrence failure is not altogether surprising given that two thirds of our sample had a prior conviction. Nor is it surprising in light of previous findings using a similar research design. Our results, based on defendants arrested on drug-related charges, are consistent with Berube and Green's (2007) results based on their study of those arrested on property-related crimes during the late 1970s and early 1980s. They also comport with Lee and McCrary's (2005) regression discontinuity analysis of Florida arrest records, which indicates that the arrival of one's eighteenth birthday dramatically increases the average sentence one is likely to receive but has no effect on one's likelihood of arrest. They are consistent as well with evidence suggesting that traditional sanctions are less effective at preventing recidivism than treatment options mandated by drug courts (Wilson, Mitchell, and MacKenzie, 2006). The experimental findings most at odds with those presented here are the specific deterrence effects that Sherman and Berk (1984) found in the wake of a randomly assigned intervention, whereby certain suspects in domestic violence cases were arrested. It might be that the contrasting results reflect the different deterrent effects of arrest as opposed to a sentence conditional on arrest and conviction. The deterrent effect observed by Sherman and Berk, which has received mixed support from follow-up experiments (Farrington and Welsh, 2005), also might be contingent on individual characteristics and social pressures at work in the context of domestic violence but absent from the world of drug crime.

The second conclusion is that probation does not alter the probability of recidivism. Again, the net effect of zero might conceal countervailing effects because the deterrent effects of probation are offset by other consequences that increase the odds of reoffending. One possibility is that supervision is not criminogenic per se; it merely increases the probability that a defendant's criminal conduct will be detected. A second possibility is that supervision is criminogenic in a limited sense; by placing additional legal constraints on a defendant, it gives him additional opportunities to violate the law. A third possibility is that supervision is criminogenic in a broader sense. Sherman (1993) has argued that the experience of continual government supervision might induce defendants to commit additional crimes in defiance of societal force.

A variety of testable propositions might help sort out whether countervailing forces account for the weak net effect of probation. If the mechanism at work is increased detection of crime, then the magnitude of this effect should vary depending on the degree of supervision that jurisdictions impose on defendants. The same should hold true under the second hypothesis in which the addition of legal constraints is the operative force. The limited experimental evidence from the United States on this question suggests that the intensiveness of supervision does not affect the rate of rearrest (Deschenes, Turner, and Greenwood, 1995; Turner, Petersilia, and Deschenes, 1992).[16] If, however, something about probation per se emboldens or provokes defendants to engage in criminal activity, then we should see a probation effect even in jurisdictions where supervision is minimal. Although the latter hypothesis might seem improbable, it is interesting to note Holden's (1983) experimental findings concerning drunk driving in Tennessee, which suggest that supervision or education programs tend to increase the rate of subsequent drunk-driving arrests.

As noted at the outset, incarceration for drug crimes occurs on a massive scale in the United States. Proponents of severe punishment for drug-related crimes rely on several arguments, one being the specific deterrence theory. This theory holds that incarceration impresses on the defendant the seriousness of the offense and the likely costs of any future infractions. Our data provide little support for this theory. Those assigned by chance to receive prison time and their counterparts who received no prison time were rearrested at similar rates over a 4-year time frame.

It seems that the defendants in our sample were unresponsive to the severity of punishment that they personally received; if anything, our evidence hints that punishment net of incapacitation increased their probability of reoffending. That this aversive experience did not diminish their subsequent criminal conduct might reflect myopic reasoning (Lee and McCrary, 2005) or the overwhelming influence of contextual factors (Kubrin and Stewart, 2006).

The question going forward is how readily these results generalize to other settings, crimes, and defendants. One could argue that drug offenders in the District of Columbia have many of the same attributes as drug offenders elsewhere (cf. Centers and Weist, 1998), but given the data at hand, we can only speculate about the generalizability of our findings to other urban settings and types of offenders. However, the research paradigm used here may be applied to any jurisdiction that assigns defendants to judges either at random or using a deterministic process not subject to

16. Observational evidence seems to support the view that probation reduces recidivism (Petersilia, 1997: 187), which underscores the need for more experimental investigation.

discretion or self-selection. Many such jurisdictions exist in the United States, and it is only a matter of time before this line of research is extended to new sources of data that will speak to the external validity of our findings. Larger quantities of data also would support a more finegrained analysis of the interaction between sanctions and background factors such as criminal history and socioeconomic conditions.

Should the results obtained here hold up, the behavioral and policy implications are profound. From a behavioral standpoint, the ineffectiveness of punishment for subsequent criminal conduct speaks to long-standing questions about whether human behavior can be changed through aversive conditioning. From a policy standpoint, ineffective specific deterrence means that punitive policies must be justified by other considerations, such as general deterrence, retribution, or incapacitation.

REFERENCES

Achen, Christopher H. 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.

Angrist, Joshua D., Guido W. Imbens, and Donald D. Rubin. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91:444–55.

Angrist, Joshua D., and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Arceneaux, Kevin. 2005. Using cluster randomized field experiments to study voting behavior: The science of voter mobilization. *The Annals of the American Academy of Political and Social Science* 601:169–79.

Berube, Danton, and Donald P. Green. 2007. The effects of sentencing on recidivism: Results from a natural experiment. Paper presented at the Second Annual Conference on Empirical Legal Studies, New York.

Blumstein, Alfred, Jacqueline Cohen, and Daniel S. Nagin, eds. 1978. *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*. Washington, DC: National Academy Press.

Blumstein, Alfred, and Richard C. Larson. 1971. Problems in modeling and measuring recidivism. *Journal of Research in Crime and Delinquency* 8:124–32.

Bound, John, David A. Jaeger, and Regina Baker. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association* 90:443–50.

Centers, Nathan L., and Mark D. Weist. 1998. Inner city youth and drug dealing: A review of the problem. *Journal of Youth and Adolescence* 27:395–411.

Chernozhukov, Victor, and Christian Hansen. 2008. The reduced form: A simple approach to inference with weak instruments. *Economic Letters* 100:68–71.

Deschenes, Elizabeth P., Susan Turner, and Peter W. Greenwood. 1995. Drug court or probation? An experimental evaluation of Maricopa County's drug court. *The Justice System Journal* 18:55–73.

District of Columbia Sentencing Commission. 2000. *D.C. Act 13-406, the 'Sentencing Reform Amendment Act of 2000.'* http://acs.dc.gov/acs/lib/acs/pdf/acs.DCAct13-406.pdf.

Farrington, David P., and Brandon C. Welsh. 2005. Randomized experiments in criminology: What have we learned in the last two decades? *Journal of Experimental Criminology* 1:9–38.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Gottfredson, Don M. 1999. *Effects of Judges' Sentencing Decisions on Criminal Careers*. Washington, DC: National Institute of Justice.

Hansen, Ben B., and Jake Bowers. 2008. Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science* 23:219–36.

Hausman, Jerry A. 1978. Specification tests in econometrics. *Econometrica* 46:1251–71.

Holden, Robert T. 1983. Rehabilitative sanctions for drunk driving: An experimental evaluation. *Journal of Research in Crime and Delinquency* 20:55–72.

Imbens, Guido W., and Joshua D. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62:467–75.

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

Killias, Martin, and Patrice Villettaz. 2008. The effects of custodial vs non-custodial sanctions on reoffending: Lessons from a systematic review. *Psicothema* 20:29–34.

Kim, Il-Joong. 1994. An econometric study on the deterrent impact of probation. *Evaluation Review* 18:389–410.

Kim, Il-Joong, Bruce L. Benson, David W. Rasmussen, and Thomas W. Zuehlke. 1993. An economic analysis of recidivism among drug offenders. *Southern Economic Journal* 60:169–83.

Kling, Jeffrey R. 2006. Incarceration length, employment, and earnings. *American Economic Review* 96:863–76.

Kubrin, Charis E., and Eric A. Stewart. 2006. Predicting who reoffends: The neglected role of neighborhood context in recidivism studies. *Criminology* 44:165–97.

Lacasse, Chantale, and A. Abigail Payne. 1999. Federal sentencing guidelines and mandatory minimum sentences: Do defendants bargain in the shadow of the judge? *Journal of Law and Economics* 42:245–69.

Landes, William M. 1971. An economic analysis of the courts. *Journal of Law and Economics* 14:61–107.

Lee, David, and Justin McCrary. 2005. Crime, punishment, and myopia. NBER Working Paper No. W11491.

Maltz, Michael D., and Richard McCleary. 1977. The mathematics of behavioral change: Recidivism and construct validity. *Evaluation Quarterly* 1:421–37.

Manski, Charles F., and Daniel S. Nagin. 1998. Bounding disagreements about treatment effects: A case study of sentencing and recidivism. *Sociological Methodology* 28:99–137.

Martin, Susan E., Sampson Annan, and Brian Forst. 1993. The special deterrent effects of a jail sanction on first-time drunk drivers: A quasi-experimental study. *Accident Analysis and Prevention* 25:561–8.

Massoglia, Michael, and Ross Macmillan. 2002. Rational choice, deterrence, and criminal offending: A consideration of legal subjectivity. In *Rational Choice and Criminal Behavior: Recent Research and Future Challenges*, eds. Alex R. Piquero and Stephen G. Tibbetts. New York: Routledge Press.

Murakawa, Naomi. 2005. Electing to punish: Congress, race, and the American criminal justice state. PhD Dissertation. Yale University, New Haven, CT.

Nagin, Daniel S., Francis T. Cullen, and Cheryl Lero Jonson. 2009. Imprisonment and reoffending. In *Crime and Justice: A Review of Research*, vol. 38, ed. Michael H. Tonry. Chicago, IL: University of Chicago Press.

Orsagh, Thomas, and Jong-Rong Chen. 1988. The effect of time served on recidivism: An interdisciplinary theory. *Journal of Quantitative Criminology* 4:155–71.

Persson, Mats, and Claes-Henric Siven. 2006. Incentive and incarceration effects in a general equilibrium model of crime. *Journal of Economic Behavior & Organization* 59:214–29.

Petersilia, Joan. 1997. Probation in the United States. In *Crime and Justice: A Review of Research*, vol. 22, ed. Michael H. Tonry. Chicago, IL: University of Chicago Press.

Robinson, Paul H., and John M. Darley. 2004. Does criminal law deter? A behavioral science investigation. *Oxford Journal of Legal Studies* 24:173–205.

Sherman, Lawrence. 1993. Defiance, deterrence, and irrelevance: A theory of the criminal sanction. *Journal of Research in Crime and Delinquency* 30:445–73.

Sherman, Lawrence, and Richard A. Berk. 1984. The specific deterrent effects of arrest for domestic assault. *American Sociological Review* 49:261–72.

Smith, Douglas A., and Patrick R. Gartin. 1989. Specifying specific deterrence: The influence of arrest on future criminal activity. *American Sociological Review* 54:94–106.

Spohn, Cassia, and David Holleran. 2002. The effect of imprisonment on recidivism rates of felony offenders: A focus on drug offenders. *Criminology* 40:329–57.

Staiger, Douglas, and James H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65:557–86.

Stock, James H., and Mark W. Watson. 2007. *Introduction to Econometrics*, 2nd ed. New York: Pearson Addison-Wesley.

Stock, James H., and Motohiro Yogo. 2002. Testing for weak instruments in linear IV regression. NBER Technical Working Paper No. 284.

Turner, Susan, Joan Petersilia, and Elizabeth P. Deschenes. 1992. Evaluating intensive supervision probation/parole (ISP) for drug offenders. *Crime and Delinquency* 38:539–56.

United States General Accounting Office (USGAO). 2001. *D.C. Criminal Justice System: Better Coordination Needed Among Participating Agencies*. GAO Pub. No. 01-187. Washington, DC: GPO. http://www.gao.gov/new.items/d01187.pdf.

Villettaz, Patrice, Martin Killias, and Isabel Zoder. 2006. *The effects of custodial vs. non-custodial sentences on reoffending: A systematic review of the state of knowledge*. Report to the Campbell Collaboration Crime and Justice Group. http://www.campbellcollaboration.org/doc-pdf/Campbell-report-30.09.06.pdf.

Waldfogel, Joel. 1994. Does conviction have a persistent effect on income and employment? *International Review of Law and Economics* 14:103–19.

West, Heather C., and William J. Sabol. 2008. *Prisoners in 2007*. NCJ Pub. No. 224280. Washington, DC: Bureau of Justice Statistics. http://www.ojp.usdoj.gov/bjs/pub/pdf/p07.pdf.

Wilson, David B., Ojmarrh Mitchell, and Doris L. MacKenzie. 2006. A systematic review of drug court effects on recidivism. *Journal of Experimental Criminology* 2:459–87.

Witte, Ann D. 1980. Estimating the economic model of crime with individual data. *Quarterly Journal of Economics* 94:57–84.

Donald P. Green is A. Whitney Griswold Professor of Political Science at Yale University and director of Yale's Institution for Social and Policy Studies. His research interests include hate crime, rationality, and experimental research methods.

Daniel Winik is a student at Yale Law School, where he serves as an editor of the *Yale Law Journal* and the *Yale Law & Policy Review*. His main academic interest is in public law.