J Quant Criminol (2007) 23:303–326
DOI 10.1007/s10940-007-9032-4

ORIGINAL PAPER

Incapacitation: Revisiting an Old Question with a New Method and New Data

Gary Sweeten · Robert Apel

Published online: 10 July 2007

Abstract We use the National Longitudinal Survey of Youth 1997 to obtain estimates of the number of crimes avoided through incapacitation of individual offenders. Incarcerated individuals are matched to comparable non-incarcerated counterparts using propensity score matching. Propensity scores for incarceration are calculated using a wide variety of time-stable and time-varying confounding variables. We separately analyze juvenile (age 16 or 17) and adult (age 18 or 19) incapacitation effects. Our best estimate is that between 6.2 and 14.1 offenses are prevented per year of juvenile incarceration, and 4.9 to 8.4 offenses are prevented per year of adult incarceration.

Keywords Incapacitation · Incarceration · Propensity score matching · Juvenile justice · Prison

A daunting challenge for any study of incapacitation is to estimate what would have happened had the incarcerated individual been free. In other words, on average, how many crimes are prevented by incarcerating an individual? The counterfactual outcome—how many crimes the person would have committed if free—is unobserved and must be inferred. If the justice system is operating as it should, those incarcerated are the most serious and prolific offenders in the population. The challenge, therefore, is to identify a comparable non-incarcerated group that can be used to estimate a valid incapacitation effect on criminal offending. Barring observable exogenous variation in the incarceration decision, this challenge must be met with non-experimental analysis.

G. Sweeten (✉)
School of Criminology and Criminal Justice, Arizona State University, Mail code 3250, PO Box 37100, Phoenix, AZ 85069-7100, USA
e-mail: Gary.Sweeten@asu.edu

R. Apel
School of Criminal Justice, University at Albany-SUNY, Albany, NY, USA

The underlying issue is one of *balance*, or ensuring that "treated" (incarcerated) and "untreated" (non-incarcerated) individuals are statistically equivalent on all background factors relevant for estimating incapacitation effects. The goal is to simulate the conditions of random assignment to incarceration so that valid behavioral comparisons can be made. To the extent that we can achieve balance on potential confounders, we can be confident that we have approximated the conditions of a randomized experiment. This problem has received a considerable amount of attention in the evaluation literature, and a variety of techniques have become available to address it (Heckman and Hotz 1989). In our study, we employ an approach that entails matching incarcerated individuals with non-incarcerated individuals on the basis of a propensity score created using information temporally prior to the treated individual's first incarceration spell.

A Bottom-Up Approach to Estimating Incapacitation Effects

Spelman (2000) identifies two general kinds of prison effectiveness research. "Top-down" approaches use aggregate data to estimate the effect of prison population or prison commitments on crime rates. These studies are, by necessity, agnostic about the specific mechanism by which prisons influence behavior (i.e., through incapacitation, deterrence, or rehabilitation), but are often preferred by policy analysts for their ability to estimate the *total effect* of imprisonment.[1] This research tradition commenced with Gibbs (1968) and included such early and widely cited studies as Tittle (1969) and Ehrlich (1973), to name but a few. This tradition continues to more recent times in such notable and more analytically sophisticated works as Marvell and Moody (1994), Levitt (1996), and Kovandzic and Vieraitis (2006).

In this study, we adopt a second, "bottom-up" approach to studying prison effectiveness. Bottom-up approaches, according to Spelman (2000), attempt to peer into the black box of prison effectiveness using individual-level data to identify prison effects. These studies typically address the incapacitation question by using arrestee or inmate samples to obtain self-report estimates of their offending frequency in the months prior to criminal justice intervention (referred to in the criminal career literature as "lambda" or λ; see Blumstein et al. 1986), justice system processing probabilities conditional on crime commission, and average sentence lengths. These estimates are then inserted into established mathematical models to estimate the elasticity of crime with respect to incarceration (Avi-Itzhak and Shinnar 1973). Notable studies in this tradition include Shinnar and Shinnar (1975), Greenwood and Abrahamse (1982), and Horney and Marshall (1991).

Spelman (2000) critiques bottom-up studies on the grounds that they typically underestimate the total effects of prison expansion (by ignoring deterrence and rehabilitation), impose often unrealistic assumptions on model parameters (this is at least true of earlier model formulations), neglect juvenile offending (which accounts for a sizable portion of crime), and suffer from intolerable imprecision. We would add as well that existing bottom-up studies have typically employed samples of incarcerated offenders with questionable

[1] It bears mentioning that these studies still generally purport to estimate the deterrent effect of imprisonment. However, Nagin (1978) rightly points out that these studies actually confound deterrence and incapacitation (not to mention rehabilitation), and Spelman (2000) acknowledges that by treating the crime prevention mechanism as a "black box," top-down studies are unable to discriminate between competing explanations for prison effectiveness.

generalizability to contemporary circumstances (see Zimring and Hawkins 1995, for a similar critique). Prior individual-level studies base incapacitation estimates on inmates who were incarcerated near the beginning of an unprecedented expansion in U.S. prisons. From the 1930s through the early 1970s, the incarceration rate hovered around 110 per 100,000 residents, then began a steady increase in the early 1970s and at midyear 2005 had attained 738 per 100,000 residents (Harrison and Beck 2006). If the justice system operates efficiently, it will identify and incarcerate the most serious offenders—those who commit the most serious crimes and who do so at a high rate. Holding all else equal, however, rapid expansion in the prison population will also result in less active or less serious offenders entering prison. Contemporary incarceration thus might yield lower incapacitative benefits than incarceration in earlier decades because of diminishing marginal returns as the criminal justice system reaches deeper into the offender queue. Therefore, incapacitation studies conducted in the 1980s and earlier may have limited utility for policy makers today, and may in fact overestimate the current incapacitative benefit of prison.

However, despite their shortcomings, we believe that it is premature to dismiss bottom-up studies altogether. It is worth observing that there are multiple strategies to estimate incapacitation effects. Zimring and Hawkins (1995: 81) make this point succinctly:

> [T]o determine the level of crime that would have occurred if a particular group had not been confined, one must either study the criminal activity of the same group at a different time in their lives to estimate what that group would have done if not confined, or one must study the behavior of persons other than those confined to approximate the crimes avoided by imprisonment in the past.

We label these within-individual and between-individual strategies, respectively. Existing bottom-up studies of incapacitation are of the former type, in which targeted individuals are surveyed about their criminal activity during time periods when they were on the street rather than confined. These estimates are then taken as lambda, or the number of crimes they committed on an annual basis when they were free, and by implication, the number of crimes they *would have committed* per year had they not been in jail or prison. Thus, the typical counterfactual offense rate for incarcerated individuals is their own offense rate in the months preceding their confinement. However, one of the most pertinent shortcomings of existing within-individual, bottom-up studies is the question of the reliability and validity of offending rates reported by arrested or incarcerated individuals (Spelman 1994; Zimring and Hawkins 1995). For one, it is unclear whether incarcerated offenders, particularly high-rate offenders, can accurately recall their prior criminal activity. Even if they can, they may not be so inclined to report it honestly to interviewers. Spelman (1994), in fact, finds that both overreporting and underreporting are present in the second Rand inmate study, but at different locations in the distribution of offending rates. An additional complexity is the presence and nature of crime spurts, which Blumstein et al. (1986) explain can introduce serious distortions in estimates of offending frequency. If offenders experience relatively short periods of high-rate offending immediately prior to incarceration, offense rates will be overstated, especially if reporting windows are comparatively narrow (Rolph et al. 1981).

Unfortunately, there is a conspicuous absence of between-individual, bottom-up incapacitation studies, no doubt due to the increased difficulty in recruiting and surveying a "non-captive" population. These kinds of studies may be well suited to overcoming some of the limitations of within-individual, bottom-up studies identified above. The goal of this strategy is to identify a non-incarcerated sample such that it is comparable to the incarcerated sample with the singular exception that it does not receive the "treatment" of interest. The counterfactual offense rate can thus be viewed as the offense rate of a comparison sample of non-incarcerated individuals during the time that the incarcerated sample is confined. Ideally, untreated individuals will exhibit a criminal and arrest history similar to treated individuals, but are dissimilar in that they remain in the community rather than an institution.

In this study, we attempt to generate precise estimates of one piece of the analytical puzzle—offending frequency per year of street time (the "offending rate")—while avoiding some of the pitfalls of prior research in the bottom-up tradition. First, in contrast to most prior studies, which rely on self-report offending among arrestee or inmate samples, we use a national probability sample of young people with self-report data on criminal behavior. Using self-report data, we are able to observe behavior that fails to come to the attention of criminal justice authorities. Additionally, because our sample is young and nationally representative, we can focus on populations that are at high risk for criminal activity and incarceration. Second, we use longitudinal data to identify incarceration risk prospectively, which will allow us to quantify the sample selection biases that are likely to be present in existing arrestee and inmate studies. Furthermore, because we focus on the first spell of incarceration, we can be sure that our estimate of the prison effect (if any) is measuring incapacitation rather than deterrence or rehabilitation. Finally, because we have detailed information on justice system processing, family background, schooling, socioeconomic indicators, and numerous other domains, we are able to approach the question of offending frequency from a novel empirical angle. Existing studies rely exclusively on within-individual variation for estimates of offending frequency. Because we have information available on non-incarcerated individuals, we employ matching methods to identify those non-incarcerated individuals who most closely resemble the incarcerated sample and use these to generate alternate estimates of offenses prevented through incapacitation.

Data

We use data from the first six waves of the National Longitudinal Survey of Youth 1997 (NLSY97) for this study. The NLSY97 is a nationally representative sample of 8,984 youths born during the years 1980 through 1984 and living in the United States during the initial interview year in 1997. The obvious disadvantage to a probability sample is the fact that comparatively few individuals may have experienced the extreme "event" of incarceration. However, roughly 5% (n = 453) of the NLSY97 sample experience at least one spell of incarceration in a jail, adult prison, or juvenile institution during the first six waves of the survey (1997–2002). To examine incarceration prospectively, we exclude from analysis the 58 individuals who were ever institutionalized prior to the initial interview or whose incarceration history is missing at this wave. Of the remaining 8,926 individuals, 396 (4.4%) experience a spell of incarceration in a jail, adult prison, or juvenile institution between the second wave (1998) and sixth wave (2002). Of these, 262 (66.2%) are incarcerated between the ages of 16 and 19, and it is this subsample that forms the basis for our propensity score analysis.

Variables

The NLSY97 contains self-report information on criminal behavior and criminal justice involvement. Each year, individuals report participation and frequency of offending in the following six activities: Intentional destruction of property, petty theft (under 50 dollars), major theft (over 50 dollars, including automobile theft), other property crimes (e.g., fencing stolen goods), attacking someone with the intent to commit serious harm, and selling illegal drugs. Individuals can report up to 99 instances of each of these six activities. We then construct a self-report offending rate per year by dividing the total number of reported offenses by the length (in years) of the reference window.[2] This self-report offending rate is used as the dependent variable in all analyses. Individuals also report the number of arrests for criminal behavior that they have experienced, and how far in the justice system they advanced for each arrest: Charged, appeared in court, convicted, incarcerated, or probated.

The NLSY97 also contains a rich array of time-stable and time-varying characteristics that we use to estimate propensity scores and assess balance for matching models. These variables include sex, race, ethnicity, urban or rural residence, condition of residence, region, type of living arrangement (house, apartment, or other), victimization, school misbehavior, attachment to school, academic achievement, gang involvement, anti-social peers, sexual activity, substance use, parental background, attachment to parents, and household income. An exhaustive list of these variables is provided in the appendix.

Methods

As noted, the challenge of any incapacitation study is to estimate a counterfactual for incarcerated individuals, where the counterfactual represents a plausible projection of the number of crimes they would have committed if they were in the community rather than an institution. Our goal in this paper is to estimate the number of crimes avoided through incapacitation among those incarcerated for the first time at two different age ranges: 16–17 and 18–19. This will allow us to contrast the incapacitation effect for juveniles versus adults.

Propensity Score Matching

Our strategy represents a "selection on observables" approach to the estimation of the incapacitation effect (see Heckman and Hotz 1989). We use observed individual characteristics to construct a propensity score for incarceration, defined as "the conditional probability of assignment to a particular treatment given a vector of observed covariates" (Rosenbaum and Rubin 1984: 516; see also Rosenbaum and Rubin 1983). We write the propensity score, $e(x)$, in the following way:

$$e(x) = P(Incarceration = 1|X)$$

where *Incarceration* indicates the treatment assignment (0,1) and X represents a vector of observed covariates that are presumed to be correlated with either incarceration status or offending. For our purposes "treatment" consists of any length of incarceration during the particular age range of interest. In this analysis, we use the cumulative logistic function

[2] When we restrict our attention to the incarcerated subsample, the length of the reference window is reduced by the length of confinement during the wave of first incarceration.

with a theoretically relevant set of prospective predictors to estimate the propensity score (see Appendix).

While propensity scores are generated using a logistic regression model, propensity score matching is essentially a non-parametric comparison of treated and untreated cases. Propensity score matching relies on fewer assumptions than standard regression estimates, and it can reveal instances where standard regression models are inappropriate. For instance, in situations where there are many unmatched cases, it becomes clear that regression-based procedures use off-support extrapolation to generate treatment effect estimates. These benefits are not without cost. In situations with highly skewed dependent variables (such as self-report offending rates), matching estimates can be particularly imprecise.

The goal of propensity score matching is to balance the observed covariates between the incarcerated and non-incarcerated individuals, conditional on the propensity score $e(x)$. If this goal is met, treatment is assumed to be random conditional on the propensity score. This is known as the conditional independence assumption (CIA). Evidence for the CIA is assessed using a measure of standardized bias that compares covariates among the treated and matched untreated individuals. This measure, first described by Rosenbaum and Rubin (1985: 36), begins with calculation of *unadjusted bias*, which is the difference between the treated and untreated on a particular characteristic divided by an equally weighted combination of the standard error within the two groups (and multiplied by 100). If this statistic exceeds 20, the characteristic is considered unbalanced. *Adjusted bias* is calculated in the same way except the treated are compared to the matched untreated cases. The standard error remains the same. If the matched sample reduces bias below 20, the covariate is considered balanced.

The conditional independence assumption, and evidence in support of it, is extremely important for the validity of matching estimates. A standard criticism of propensity score matching, and all selection-on-observables methods, is that some unobserved factor accounts for the differences between the two groups. While this challenge can never be ignored, the response relies on the conditional independence assumption. If the conditional independence assumption is valid, there is no unobserved factor that differs between the two groups. The CIA also protects against the criticism of regression to two different means accounting for the observed treatment effect (Campbell and Erlebacher 1970). If the CIA is met, treated and untreated individuals with the same propensity score are equally likely to receive treatment, and are thus drawn from the same distribution.[3] Because of the importance of the CIA, we compare our treated and untreated cases on the basis of many more factors (over 100) than are used to estimate the propensity score. To the extent that the propensity score balances background predictors, including those not included in the propensity score estimation, the CIA is given greater support.

Once propensity scores are obtained, there are a number of methods for matching untreated to treated cases. The simplest is *nearest neighbor matching*, in which the untreated case with the closest propensity score to a treated case is used as a comparison (Smith and Todd 2005). There are several variants to the method: Matching can be done with or without replacement, and individuals can be matched to one or several of their nearest neighbors within a certain range. *Kernel matching* weights untreated cases according to their distance from treated cases on the propensity score metric. In fact, all matching methods may be characterized as weighting functions, but kernel matching allows for finer distinctions in weighting than other methods. As with nearest neighbor

[3] We thank an anonymous reviewer for bringing these issues to our attention.

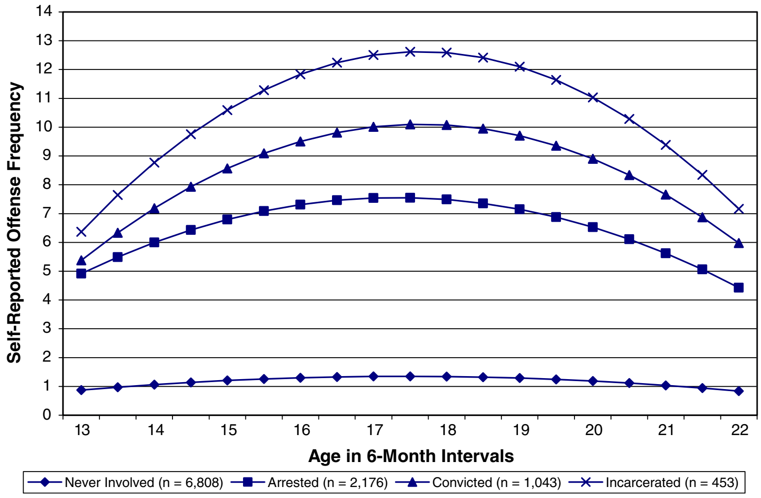

Fig. 1 Age distribution of self-report offenses in the NLSY97, by level of criminal justice involvement *Note*: The "Never Involved" and "Arrested" subsamples are mutually exclusive. The "Arrested," "Convicted," and "Incarcerated" subsamples are not mutually exclusive. The *n*s represent the number of NLSY97 respondents who advance at least as far into the criminal justice system during the first six waves

matching, there are numerous variations to kernel matching. The choice of kernel had little effect on our final estimates, so we employed the commonly used Epanechnikov kernel (Epanechnikov 1969).[4]

We should point out that the logic of matching can also be applied without the estimation of propensity scores. Any characteristic may be used to match treated and untreated cases. The effectiveness of simple matches using a few characteristics is assessed with the adjusted bias measure, as with propensity score matching. Comparing the percentage of balanced characteristics across different matching methods allows assessment of the extent to which each method achieves conditional independence. In this study, we will employ three different matching methods: Simple matching based on contact with the justice system, nearest neighbor matching, and Epanechnikov kernel matching. In each case, our estimate of the incapacitation effect is obtained by measuring the self-report offending rate among the matched unincarcerated sample during the age range of interest. No propensity score is estimated for simple matching. Instead, incarcerated youths are matched as a group to unincarcerated youths who experienced a certain level of justice system processing (i.e., arrest, conviction) *prior* to the treatment age of interest.

Results

We begin with a simple description of the criminal history among all 8,984 respondents in the NLSY97. In Fig. 1, we present age-crime curves for the following sub-samples: Youths with no involvement in the criminal justice system during the first six waves, youths with at least one arrest, youths with at least one conviction, and youths with at least one

[4] This was also the default kernel in Stata module psmatch2 (Leuven and Sianesi, 2003), which was used for all propensity score estimates in this study.

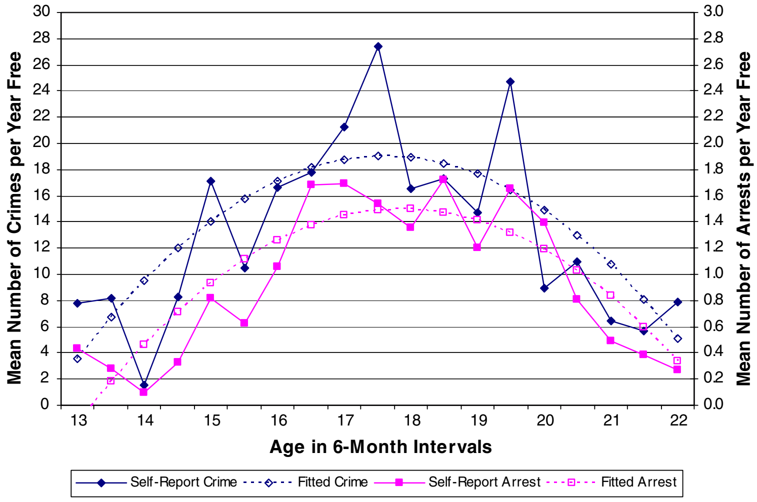

Fig. 2 Mean number of self-report offenses and arrests per year free among individuals first incarcerated between ages 16 and 19, by age. *Note*: The data in this figure are based on 1,455 person-periods for 262 individuals, and reflect the mean crime and arrest frequency per year of street time. "Fitted" figures are obtained from regressions of crime and arrest against age and its square, using all available observations during the first six waves

incarceration spell. Criminal justice involvement obviously does not arise completely at random as illustrated by the fact that youths with no involvement report a much lower level of criminal behavior. This is a clear illustration of heterogeneous offending rates and the "stochastic selectivity" that lead to high-λ individuals being filtered further into the justice system (Blumstein et al. 1993). For example, youths with at least one arrest engage in more crime by a factor approaching 5.0 at all ages compared to youths with no criminal justice involvement. In addition, the level of the age-crime curve is correlated with cumulative involvement in the criminal justice system. This is to say that youths who have ever been convicted commit crime at a much higher rate than youths who have ever been arrested, and that youths who have ever been incarcerated commit crime at a much higher rate than youths who have ever been convicted.[5] This figure thus provides a clear illustration of the problem that inheres in identifying a plausible comparison group for any form of criminal justice involvement in the absence of additional background information.

We next provide a description of the self-report offending and arrest history of our sample of 262 youths who are first incarcerated between the ages of 16 and 19. This group constitutes our estimation sample for the propensity score model. In Fig. 2, we plot the mean number of self-report crimes and arrests for this subsample from ages 13 to 22. Among this subsample, criminal behavior peaks at age 17.6 years with 19.0 crimes per year of street time, on average, or 1.6 crimes per month free. Arrest peaks at age 17.9 years with 1.5 arrests per year of street time, on average. From an early age, these to-be-incarcerated

[5] Note that Fig. 1 provides "effective" offense rates, or the number of crimes committed since the previous interview, without subtracting off the length of confinement for those individuals who are incarcerated. Fig. 1 also shows that the shape of the age distribution of crime remains unchanged despite the fact that we estimated separate age-crime curves for the subsamples. Specifically, for very different groups of individuals, we observe the usual, unimodal peak in self-report crime at about 17.5 years of age (see Hirschi and Gottfredson 1983).

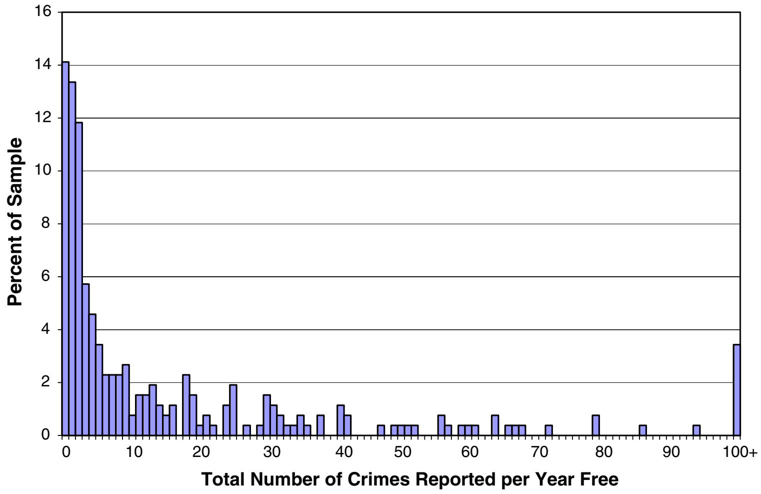

Fig. 3 Histogram of total number of self-report offenses per year free among individuals first incarcerated between ages 16 and 19. *Note*: The data in this figure are computed for 262 individuals, and reflect the total crime frequency per year of street time (lambda) in all time periods up to and including the wave in which they are first incarcerated. Estimates are rounded to the nearest integer

youths are quite precocious in their criminal behavior. For example, at age 14 they already report 9.5 crimes per year free, on average, as well as 0.5 arrests (or one arrest every 2 years). It is not until they reach their early 20s that these youths appreciably reduce their criminal involvement.[6]

Figure 3 provides an alternative representation of individual offending rates by summing offenses across all time periods and ages prior to first incarceration. This figure reveals a highly skewed distribution of lambda, or the mean number of crimes reported per year of street time. Our 262 to-be-incarcerated individuals average 18.7 crimes per year free before they are incarcerated for the first time (median = 4.6), which translates into 1.6 crimes per month free. Notably, the top 10% of this distribution of individuals report committing over 50 crimes per year (90th percentile = 50.9), or 4.2 crimes per month.[7]

[6] We can also compare the age-crime curve in Fig. 1 for the 453 individuals who have ever been incarcerated with Fig. 2 for the 262 individuals incarcerated for the first time between the ages of 16 and 19. The peak offending rate in Fig. 1 is 12.6, whereas in Fig. 2 it is 19.0, but we caution that these estimates are not exactly comparable. The difference is only partially due to the fact that Fig. 1 provides offense frequency per year since the last interview, while Fig. 2 provides offense frequency per year while on the street. It is also due to the fact that Fig. 1 includes individuals incarcerated for the first time after age 19, who we may presume have a later age of onset and thus lower criminal propensity than individuals incarcerated for the first time before age 19 (although this is mildly offset by the handful of individuals incarcerated for the first time before age 16). To the extent that age of onset is negatively correlated with criminal propensity, individuals incarcerated for the first time at a younger age should be more heavily involved in criminal behavior at any given age than individuals incarcerated for the first time at a later age, all else equal (see Gottfredson and Hirschi 1990).

[7] The information in Fig. 3 includes the wave in which respondents are incarcerated for the first time. However, in light of an absence of event history information at each time period as well as evidence of a "crime spurt" immediately prior to incarceration, as we show below, we calculate the mean offense rate while free to be 14.2 (median = 2.5) if we use only time periods prior to the wave of first incarceration for our 262 individuals.

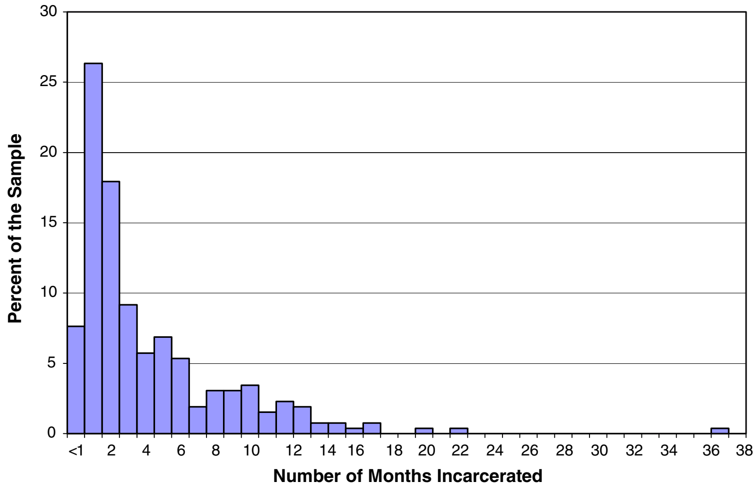

Fig. 4 Histogram of sentence length for the first incarceration spell between ages 16 and 19. *Note*: These figures are based on the 262 individuals who were first time when they were between 16 years and 19 years of age

We now turn to a description of the incarceration history of the youths in our estimation sample. Among the 262 individuals who experience at least one spell of incarceration, 113 (43.1%) do so for the first time as a juvenile (16 or 17), and 149 (56.9%) experience first-time incarceration as an adult (18 or 19). Three-quarters of these individuals (n = 193 or 73.7%) are incarcerated during only one wave, one-fifth (n = 59 or 22.5%) are incarcerated during two waves, and 4% (n = 10 or 3.8%) are incarcerated during three or four waves. Taken together, these 262 individuals contribute 344 person-periods of incarceration. The mean incarceration length is 4.5 months, with a range from less than one month to just over 3 years. However, the distribution of sentence length is highly skewed as illustrated in Fig. 4, implying that the mean is a less than ideal measure of central tendency. Accordingly, we note that the median sentence length (with 50.6 cumulative percent) is 2 months. Mean sentence length is also distorted by the fact that a handful of individuals serve multiple terms of confinement, and sentence length will undoubtedly increase with subsequent incarceration spells. If we consider only the first incarceration spell, mean sentence length is 4.2 months with median 2 months (n = 262). Among individuals with multiple incarceration spells, the mean length of the second spell is 4.1 months with median 3 months (n = 69), and mean length of the third or later spell is 7.4 months with median 7 months (n = 13).[8]

Figure 5 summarizes criminal history up to and including the wave in which these 262 youths are first incarcerated, showing observed offending as well as expected offending based on a quadratic age-crime curve, a waves-to-incarceration counter, and a dummy indicator for the wave in which they serve their first term of incarceration. This figure shows that youths are very criminally active prior to and during the wave that they are incarcerated for the first time (time t in the figure, where t represents interview waves). For example, 3–5 years prior to their first incarceration, these youths are involved in

[8] The mean age (at interview) of incarceration across all spells is 18.9 years. The mean age of the first spell of incarceration is 18.8 years, of the second spell is 19.5 years, and of the third or later spell is 20.4 years.

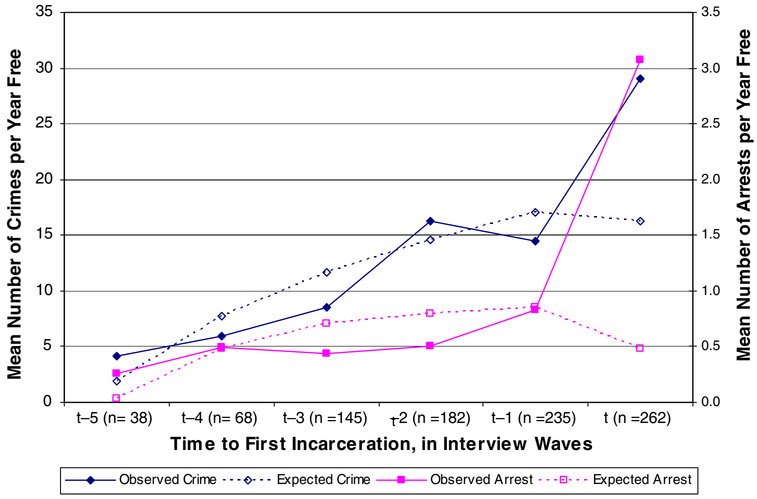

Fig. 5 Mean number of crimes and arrests per year free in time periods leading up to and including first incarceration between ages 16 and 19. *Note*: These figures are based on available data for the 262 individuals who were incarcerated for the first time when they were between the ages of 16 and 19. Time *t* represents the wave in which all individuals are incarcerated for the first time. The *n*s represent the number of individuals that contribute information during each period. Expected crime and arrest are estimated using the coefficients from a quadratic age-crime or age-arrest curve, including a time-to-incarceration counter and a dummy variable for the time period in which respondents are first incarcerated (time *t*)

approximately 7.2 crimes per year. One to 2 years before they are incarcerated, they double their criminal activity to about 15.2 crimes per year. During the wave in which they are incarcerated, they again double their criminal involvement to 29.1 crimes per year that they are free, on average. The conclusions are the same if we consider the mean number of arrests, as shown in the same figure. The conclusions are also the same if we distinguish first-time incarceration between ages 16 and 17, and first-time incarceration between ages 18 and 19 (not shown). The implication is that there is escalation of criminal activity contemporaneous with incarceration.[9] This is consistent with the observation that individuals often initiate a short period of offending at a much higher (than average) rate just prior to the arrest that leads to their imprisonment, what criminal career researchers have

[9] While this evidence suggests that individuals may be more actively involved in crime just prior to incarceration, there are two important caveats which must be noted. First, because our criterion is self-report offending, there is no reason that individuals cannot report crimes even when incarcerated during the entire reporting window at time *t*. Youths can certainly steal and get into fights while incarcerated and there is no reason they should not report this activity because it happened while they were confined. In fact, for the 15 individuals who were locked up for the entire reporting window, the reported offending rate was 10.4 crimes per year. However, the fact that the pattern is identical for self-report arrest rates prior to incarceration helps mitigate this limitation. In addition, if we replicate Fig. 5 showing criminal activity for individuals leading up to their *first arrest*, only a fraction of whom (about 21%) will subsequently be incarcerated, the same pattern emerges. Second, within a reporting window, while we are able to determine the dates of confinement, we are unable to determine the dates of reported offending. The absence of crime event history information thus means that we are unable to identify whether self-report crimes occurred before, during, or after periods of confinement. The significance of the crime spurt hinges on which of these periods reported offending falls into.

referred to as a "crime spurt" (see Blumstein et al. 1986; Rolph and Chaiken 1987).[10] In other words, lambda does not appear to be constant within individuals, thus incarceration may partly be due to elevated exposure to the risk of arrest and confinement.[11] We return to this issue for further discussion following the propensity score results.

Traditionally, incapacitation estimates are based on some version of Fig. 5. If we were to treat this as a within-individual, bottom-up incapacitation study, we would conclude that our estimate of lambda is quite sensitive to the time period of choice. Were we to choose time t in Fig. 5 as the reference point, we would be led to conclude that lambda was 29.1. That is, 1 year of incarceration per offender would prevent 29.1 crimes in the community or, since mean length of confinement is 4.5 months, 10.9 crimes for the average offender ($[29.1/12] \times 4.5$). However, it is possible that estimating offending rates from the crime spurt could lead to a gross overestimation of incapacitation effects, especially if the spurt is ephemeral (or, as we discuss later, artifactual). If, in light of the apparent crime spurt in Fig. 5, we were to choose instead time $t - 1$ as the reference point, we would conclude that lambda was 15.2. Finally, if we were to choose all available pre-incarceration time periods as our reference, as in Fig. 3, we would be led to conclude that lambda is either 18.7 or 14.2 depending on whether we include the wave of first incarceration in our computation or not.

Propensity Score Estimates of the Incapacitation Effect

We now turn to our between-individual, bottom-up estimates of incapacitation. Our first task is to assess the extent of imbalance between incarcerated and non-incarcerated individuals. We first assess differences between those who report first-time incarceration at age 16 or 17 ($n = 113$) and those who report no incarceration through age 17 ($n = 6{,}100$). The overall sample size is reduced for this comparison because we require at least one pre-16 observation in order to measure prospective background characteristics, and about 20% of the NLSY97 sample was 16 or older during the first wave of data collection. We compare the two groups on a set of 110 time-stable (measured at wave 1) and time-varying (measured during the wave prior to age 16) characteristics. Details on balance are shown in the appendix. Unadjusted bias exceeds the imbalance threshold for 73% of the assessed characteristics. Similarly, for first-time incarceration at age 18 or 19 ($n = 149$) versus those with no incarceration history through age 19 ($n = 8{,}034$), 65% of the observed characteristics are imbalanced prior to matching. The lack of balance that we observe on over 100 time-stable and time-varying characteristics between incarcerated and non-incarcerated individuals clearly demonstrates the strong selection process that is at work. The 10 least balanced variables are listed in Table 1 for incarcerated 16–17 year olds and incarcerated 18–19 year olds, along with reductions in standardized bias obtained with the various

[10] To ensure that our finding of a crime spurt in time t in these data was not due to our computing the offending rate per year free (which shortens the reference window for the incarcerated sample by subtracting off the time during which they are confined), we computed the mean offending rate during the time since the last interview (using the entire reference window rather than street time only), or what is known as the "effective" offending rate in the criminal career literature. The pattern in Fig. 5 was replicated, with lambda at time t equal to 23.7.

[11] We find that crime prevalence (an indicator for having committed *any* crime) and crime variety (the number of *different* crime types committed) increase in a fashion similar to crime frequency. Thus, incarceration appears to arise as a result of more frequent and widespread criminal involvement. We also find, interestingly enough, that although all crime types increase jointly, frequency of drug selling increases at a somewhat faster rate than frequency of other crime types in our data (e.g., vandalism, theft, assault). Thus incarceration may also be due in part to a modest change in crime mix favoring drug-related offenses.

Table 1 Bias reduction from matching methods for the ten least balanced variables

Variable (range)	Unadjusted means and standardized bias			Percent bias reduction by matching protocol			
	Incarcerated?			Simple group matching		Propensity score matching	
	Yes	No	Bias	Arrested	Convicted	Nearest neighbor	Kernel
First incarceration between ages 16 and 17							
Suspended (0/1)	0.70	0.19	119.1	65.0	86.6	98.2	98.7
Fought at school, wave 1 (0/1)	0.37	0.16	96.2	41.5	50.0	91.3	96.7
Years sexually active (0–11)	2.50	0.69	93.8	29.2	67.0	76.7	90.2
Delinquency variety (0–6)	1.94	0.53	93.5	72.9	80.5	95.9	90.9
Smoked (0/1)	0.74	0.38	77.2	99.7	82.8	100.0	87.8
Cigarettes/day (0–60)	5.45	1.02	76.3	56.7	92.1	48.9	89.9
No. of days smoked (0–30)	11.95	3.23	76.3	41.9	67.4	66.4	94.1
Low middle school grades (0/1)	0.56	0.23	72.1	68.1	79.5	94.3	87.9
Used marijuana (0/1)	0.52	0.20	71.9	87.9	52.2	85.3	92.9
Ever in a gang (0/1)	0.11	0.02	64.4	44.7	99.4	79.9	91.0
First incarceration between ages 18 and 19							
Years sexually active (0–13)	3.77	1.73	93.1	41.4	47.2	95.7	99.0
Male (0/1)	0.87	0.50	86.0	42.9	44.4	92.4	94.1
Smoked (0/1)	.80	0.43	81.6	88.3	99.0	90.9	98.8
Delinquency variety (0–6)	1.56	0.45	78.1	76.1	79.6	87.5	85.3
Prior charges (0–6)	.80	0.10	77.5	81.0	3.4	82.5	77.2
Cigarettes/day (0–60)	15.69	5.95	77.2	71.2	80.6	96.2	86.3
Ever went to court (0/1)	0.33	0.05	75.2	57.4	−140.3	82.0	80.2
Prior arrests (0–18)	1.89	0.25	73.9	63.1	85.0	81.3	77.8
Used marijuana (0/1)	0.62	0.28	71.2	91.4	83.8	68.5	80.6
Lives with bio. parents (0/1)	0.20	0.52	−69.7	40.3	50.5	95.6	99.7

Note: All variables are assessed from the interview wave immediately prior to age 16 or 18, unless otherwise noted

matching methods. Among the 10 least balanced variables for first incarceration as a juvenile are school suspension, school misbehavior, poor grades, sexual activity, delinquency, substance use, and gang involvement. Similar variables are imbalanced for first incarceration as an adult, with the addition of juvenile justice system involvement (short of incarceration) and the exception of school-related variables.

Simple Group Matching

That the incarcerated sample differs so dramatically from the non-incarcerated in terms of background characteristics is no revelation. Incarceration involves very strong self-selection processes. The question is, Can we match the incarcerated sample to a non-incarcerated comparison group in such a way that these antecedents are balanced? Table 2 provides these results. Our baseline incapacitation estimate, which assumes that incarcerated individuals are no different from the non-incarcerated in terms of expected offending, yields incapacitation effects of 3.0 for juveniles and 2.1 for adults. This is to say

Table 2 Estimates of crimes averted through incapacitation

Matching method	Number of untreated cases	Incapacitation effect	95% Confidence interval	% Variables imbalanced
First incarceration between ages 16 and 17				
Simple group matching				
Not incarcerated	6,100	3.0	2.7, 3.4	72.7
Arrested at 15	259	9.1	5.8, 12.3	35.5
Convicted at 15	90	12.1	6.6, 17.5	32.7
Propensity score matching				
Nearest neighbor	113	10.1	3.8, 21.0	7.2[a]
Kernel	113	9.2	6.2, 14.1	4.5[a]
First incarceration between ages 18 and 19				
Simple group matching				
Not incarcerated	8,034	2.1	2.0, 2.2	64.5
Arrested at 17	363	6.8	4.9, 8.6	39.1
Convicted at 17	144	7.3	4.2, 10.3	33.6
Propensity score matching				
Nearest neighbor	149	4.4	2.6, 9.8	16.4[b]
Kernel	149	6.3	4.9, 8.4	1.8[b]

Note: The incapacitation effect represents the mean offending rate (number of crimes per year on the street) for individuals in the untreated group. In propensity score matching, many-to-one matching of untreated cases is used, although the weights sum to the number of incarcerated individuals

[a] Six individuals in the treatment group were unmatched

[b] Four individuals in the treatment group were unmatched

that individuals who were not incarcerated reported 3.0 and 2.1 offenses per year of street time during the wave that our treated individuals were confined for the first time. Our first attempt at simple group matching uses prior justice system involvement. We match the 113 incarcerated juveniles, as a group, to 259 counterparts who were arrested during the wave prior to age 16 but not incarcerated through age 17. This comparison yields an incapacitation estimate of 9.1 offenses per year of street time, but 36% of the comparison variables remain imbalanced. Similarly, for incarceration at age 18 or 19, conditioning on arrest in the wave prior to age 18 yields an incapacitation effect of 6.8 offenses per year of street time, but nearly 40% of the comparison variables remain imbalanced. While conditioning on arrest dramatically reduces standardized bias in a number of imbalanced variables, as seen in Table 1 and the Appendix, this simple group matching protocol still falls far short of the mark for achieving balance on observable characteristics.

Because simple matching on arrest does not produce a suitable comparison sample, we match on the most serious justice system outcome short of incarceration: Court conviction. Only 90 youths were convicted prior to age 16 and not incarcerated through age 17. However, comparisons of this group to those incarcerated at 16 or 17 still yield many persistent differences. Although there continues to be considerable bias reduction among the least balanced variables, one-third of our more than 100 variables are still imbalanced. Consequently, our incapacitation estimate of 12.1 is suspect. We have no better success with the adult sample. The 144 individuals who were convicted during the wave prior to age 18 differ from the incarcerated sample on 34% of the background variables.

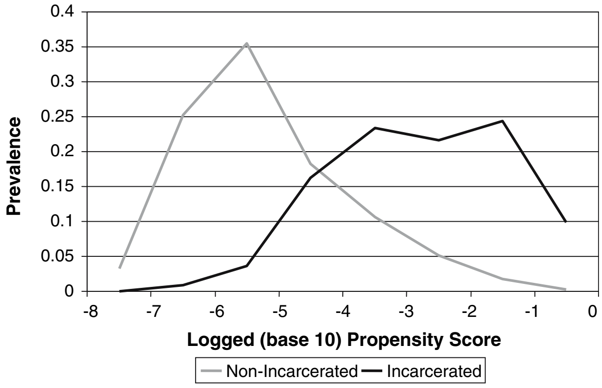

Fig. 6 Distribution of logged propensity scores by treatment status, ages 16–17. *Note*: Point estimate indicates the proportion of the subsample whose propensity score fell between the two tick marks (e.g., 3.3% of the non-incarcerated had a logged propensity score between −8 and −7, and 25.3% had a logged propensity score between −7 and −6)

Propensity Score Matching

Having little success with simple group matching, we now move to propensity score matching. For the juvenile subsample, we used a fairly concise propensity score model which yielded balance as good as any of the more complicated models we estimated. Only 23 variables were included in the propensity score model for first incarceration at 16 or 17.[12] The average propensity score for the untreated was 0.016, whereas for the treated the average was 0.133. Most propensity scores for both groups were quite low because incarceration is such a rare event. Figure 6 shows the logged (base 10) propensity score distribution for the two groups, indicating that while both groups had many low propensity scores, incarcerated individuals had considerably higher propensity scores. Propensity score matching makes transparent the comparability of treated and untreated samples. When no suitable matches are available for treated cases, it indicates that selection processes are so strong that certain configurations of background variables *always* result in treatment in the observed sample.[13] In this sample only eight individuals have propensity scores above 0.50: Six incarcerated individuals with propensity scores ranging from 0.55 to 0.85 and two non-incarcerated individuals with propensity scores between 0.65 and 0.70. Thus, there are incarcerated individuals with propensity scores below 0.65 and above 0.70 for whom no suitable matches can be found. Unmatched cases are removed from incapacitation estimates and considerations of balance. This means that our estimates are biased slightly downwards, and the generalizability of our estimates are limited accordingly.

Using nearest neighbor matching with replacement and a maximum distance between observations of 0.01 on the propensity score metric, all but six of 113 incarcerated individuals are matched to non-incarcerated counterparts. Importantly, all but 7% of the

[12] These variables (measured at the wave prior to age 16 unless noted) include: Male, black, delinquency variety, ever arrested, ever in a gang, an antisocial peer index, cigarettes smoked per day, marijuana use (0/1), suspended from school, got in fight at school (wave 1), five academic aptitude indicators (wave 1), years of sexual activity, living independently, interviewer assessment of outside of dwelling (two dummy variables), type of dwelling (two dummy variables), and residence in northeast region.

[13] This is not to say that the same configuration of background variables will always result in assignment to the treated group in all applications of this method, only that in our particular application, 100% of the youths with these configurations happen to be in the treated group. In the language of instrumental variables models for treatment effect estimation, these are analogous to the "always-takers," for whom there is no estimable counterfactual (see Angrist et al. 1996).

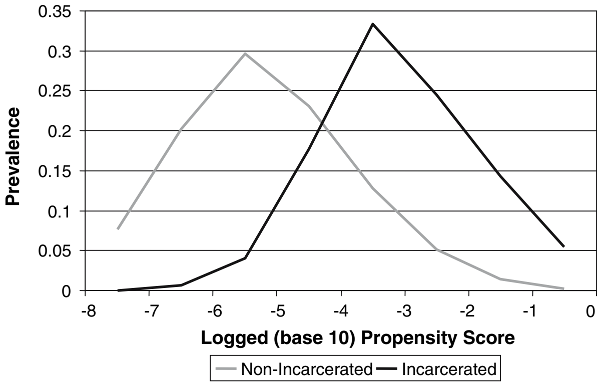

Fig. 7 Distribution of logged propensity scores by treatment status, ages 18–19. *Note*: Point estimate indicates the proportion of the subsample whose propensity score fell between the two tick marks

background characteristics are balanced using this method, and the incapacitation estimate is 10.1 with a 95% confidence interval ranging from 3.8 to 21.0.[14] Using Epanechnikov kernel matching with a bandwidth of 0.01, again, six treated cases cannot be matched. For the remaining treated and matched untreated samples, all but 5% of background characteristics are balanced, and the incapacitation estimate is 9.2. In addition, kernel matching consistently yields the greatest reduction in standardized bias among the 10 least balanced variables.[15]

An expanded propensity score model is used for the older sample, as the more concise model is unable to yield adequate balance. Additions to the model include victimization, more detailed delinquency, substance use, and justice system involvement, more parental background characteristics, and several socioeconomic status indicators for a total of 52 variables. The average propensity score was 0.015 for the untreated and 0.138 for the treated. Figure 7 displays the distributional differences in logged (base 10) propensity scores between the incarcerated and non-incarcerated individuals. Nearest neighbor matching does not perform nearly as well, even with the richer propensity score model. Four of the 149 incarcerated individuals are unmatched, and 16% of background variables are unbalanced. Epanechnikov kernel matching performs much better, with all but 2% of background variables balanced, and consistently large bias reductions in the least balanced characteristics, but still four cases are unmatched. The incapacitation estimates are 4.4 offenses per year from nearest neighbor matching and 6.3 offenses per year from kernel matching.

In sum, because they yielded the greatest balance in background factors, we regard kernel matching estimates as superior. Based on the 95% confidence interval for kernel matching, we estimate that incarcerating a juvenile for 1 year prevents between 6.2 to 14.1 offenses, and that incarcerating a young adult for 1 year prevents 4.9 to 8.4 offenses. We find it particularly interesting that our kernel matching estimates yield results that are virtually identical to a simple group matching protocol that matches incarcerated indi-

[14] With all propensity score matching estimates of incapacitation, NLSY97 sampling weights are used only in generating propensity scores. Confidence intervals are obtained from the 5th and 95th percentiles of 500 bootstrap estimates of the matching model.

[15] Males and females are included in our models, but our estimates are not gender-specific. We reproduced our propensity score matching results with a male-only sample and found substantively similar results. Three of the four matching estimates were slightly higher, and one was lower. None were statistically significantly different from reported estimates.

viduals with those who were arrested while age 15 (or age 17), but who were not incarcerated between the ages of 16 and 17 (or 18 and 19).

Discussion

The question we have addressed in this study has a fairly long history, beginning with the Rand inmate studies (Chaiken and Chaiken 1982; Peterson et al. 1980) and the subsequent search for prediction models of high-rate offenders (Greenwood and Abrahamse 1982; Greenwood and Turner 1987), and continued under the rubric of the criminal career model which expressed the individual offending rate as "lambda", λ (Blumstein et al. 1986). Several subsequent studies attempted to yield estimates of lambda through a variety of methods (e.g., Canela-Cacho et al. 1997; Horney and Marshall 1991; Nagin and Land 1993). In the 1990s, emphasis also shifted to macro-level estimates of the overall impact of incarceration, or what Spelman (2000) regards as "top-down" studies of prison effectiveness. We have taken up the pursuit of identifying the incapacitation effect using self-report offending. This study further differs from prior studies by using a contemporary, nationally representative panel dataset, and by employing propensity score matching to estimate the crime-control benefit of incarceration by way of incapacitation.

Using the technique of propensity score matching, we were able to arrive at a "best guess" estimate of the incapacitation effect of juvenile and adult prison. Our best estimate for juveniles is that 1 year of incarceration prevents 6.2 to 14.1 offenses and a year of adult incarceration prevents 4.9 to 8.4 offenses. Our incapacitation estimates are considerably smaller than those of earlier studies. The second Rand inmate study found that pre-prison offending rates were highly skewed, with a majority of offenders infrequently breaking the law and a small minority heavily involved in crime with over one hundred offenses per year (Chaiken and Chaiken 1982). Horney and Marshall (1991) used life-history calendars with an intake sample of prisoners to collect more detailed offending histories. They too found highly skewed rates of offending, with a mean offending rate of 172.2 and median of 4.2 for non-drug crimes among active offenders. Our study differs in a number of important ways from these earlier studies, so it is difficult to isolate what exactly accounts for the large difference between the estimates. We would submit, however, that the confidence intervals associated with mean offending rates in the above studies are likely quite wide, and may, in fact, be consistent with our estimates. Unfortunately, confidence intervals are not reported in these studies and cannot be calculated from the provided information.

Estimates of incapacitation are highly sensitive to the assumptions and analytical techniques applied to the question. The reason estimates are so sensitive to research design is clearly evident in Fig. 5. Viewed retrospectively, those individuals who eventually enter prison are committing crimes and accumulating arrests at a much higher rate than is typical, even for their own criminal history. It also appears that individuals are getting increasingly unlucky as their eventual incarceration approaches. For example, as shown in Fig. 5, the ratio of mean number of arrests to mean number of crimes stays fairly steady between 0.05 and 0.08, and then increases to 0.11 in the period when our sample undergoes its first spell of incarceration. Thus the crime spurt corresponds with more arrests per crime.

In economics, a similar kind of temporal selection mechanism was first noticed by Ashenfelter (1978) in an evaluation of work training programs. He noted that mean earnings of individuals who entered a government training program declined in the time period just prior to entry into the program. This finding was subsequently replicated and dubbed "Ashenfelter's Dip." The problem with limiting pre-program earnings information to the "dip" is that program impacts will be grossly overestimated.[16] Our problem is somewhat different because we are concerned with what the outcome would have been had the person *not* entered the "program," rather than estimation of post-program impacts. In criminological terms, we are interested in how much crime the person would have committed had he or she not entered prison; we are not interested, in this particular study, in post-prison (specific) deterrent effects. Also, instead of dealing with a pre-program earnings dip, we are faced with a pre-prison crime spurt (see Blumstein et al. 1986; Rolph and Chaiken 1987; Rolph et al. 1981). Regardless, the crux of the issue is the same—the degree to which there is temporal persistence of the pre-prison crime spurt. To the extent that the crime spurt is general but impermanent, the incapacitation estimate based on inmate self-reports in the period prior to the instant offense will be greatly inflated and highly inaccurate. However, if the crime spurt is a manifestation of a genuine increase in the underlying rate of offending, then incarceration can be expected to have substantial incapacitative benefits.

There is an alternative, compelling explanation for the crime spurt offered by Maltz and Pollock (1980). They point out that a rapid inflation in the number of police contacts immediately prior to sentencing to a correctional program does not necessarily mean that an individual's offending rate is truly increasing during this period. Instead, it could be attributable to the way that individuals are *selected* for correctional intervention. Specifically, judges are likely to target for incarceration individuals for whom an arrest has just occurred, and for whom a minimum number of arrests previously occurred within a fixed period of time. It so happens that such a selection rule results in an inflation of an individual's arrest rate in the months prior to intervention even though the underlying arrest rate remains constant.[17] Thus, the crime spurt prior to incarceration that we observe in our data may be more apparent than real, and thus a poor reflection of the number of crimes that are averted through incapacitation.[18]

The extent and nature of the crime spurt strikes us as an important avenue for further research in the criminal career tradition. Its existence poses a challenge for any bottom-up estimate of the effect of prison on criminal activity.[19] We hasten to add that this is true for studies of both incapacitation as well as deterrence. There are at least three ways to treat rapidly increasing offending just prior to entering prison in the present study. First, one could ignore it as a temporary anomaly, and use offending rates one or two years prior to

[16] Heckman and Smith (1999) point out that the degree to which the estimates are overestimated depends on whether the dip is permanent or transitory. If the dip is transitory and is correlated with entry into the training program, then using earnings data from the dip will severely bias estimates of program impacts.

[17] To be sure, Maltz and Pollock (1980) do not deny that individuals undergo genuine crime spurts prior to arrest and imprisonment, only that the observed crime spurt bears a strong resemblance to data simulated in such a way that an apparent crime spurt arises as a result of plausible sample selection processes.

[18] Maltz and Pollock (1980) also point out that a similar artifactual crime spurt would be observed even if individuals were endowed with two underlying, state-specific arrest rates for "active" and "quiescent" periods. Conditional on an instant arrest as the selection rule, they show that the arrest rate will decay exponentially into the past to a steady state. We are grateful to an anonymous reviewer for bringing this important article to our attention.

[19] If the crime spurt is genuine, it begs the question of what might be going on in the lives of offenders during this period of increasing volatility in their criminal history, and whether it is possible to intervene before things get out of hand and they are incarcerated.

entering prison as the appropriate offense rate estimate. Second, one could estimate the offense rate as late in the pre-prison period as possible, and take this as the incapacitation effect. Third, one could add a linear term, and assume that the person would have continued to increase his offending barring intervention from the justice system. Without experimental data, the way researchers deal with the pre-prison crime spurt must rest on their beliefs about how persistent it is. The lack of temporal resolution in our data unfortunately prevents us from choosing between these options. Because we cannot establish event history within any wave, we are limited to using self-report offending from the wave immediately prior to incarceration, forcing us to assume that the crime spurt is ephemeral. Our matching estimates are obviously affected by this assumption. We match future inmates to non-inmates based on $t - 1$ data and earlier. Although we achieve satisfactory matches at $t - 1$, we are blind to other time-dependent selection processes that occur after time $t - 1$. For example, although with kernel matching we obtain estimates of 9.2 and 6.3 crimes prevented per year using offending of the matched sample only, the incarcerated samples report 19.9 and 14.3 crimes per year, respectively, during the year of incarceration. We would predict that the smaller the temporal unit used in a study of this kind, the greater the incapacitation estimates one would derive from matching, and the more incarcerated individuals one would be unable to match.

Conclusion

There are a number of limitations of our study worth highlighting. First, the difference between our and prior incapacitation effects stems in part from our self-report offending frequency scale. This scale contains only six items. Many instances of crime and delinquency may not be covered by these six questions. Incapacitation effects would be arguably higher if a wider breadth of crime types were taken into account. Second, not all incarcerated individuals are included in our matching estimates. Six of 113 of those incarcerated as juveniles and four of 149 incarcerated as adults were not matched to non-incarcerated individuals and do not figure in our incapacitation or balance estimates. However, even if the counterfactual cases for these individuals were 10 times higher than the average counterfactual for the matched cases, it would not drive our estimates outside our estimated confidence intervals. Third, the NLSY97 has a non-trivial amount of attrition over time, as only 77% of the sample participates at each of the six interviews that we use in this analysis. Our incapacitation estimates could in fact be biased downward if individuals who drop out of the survey are not a random subset of the original sample, and if incarcerated individuals are more likely to drop out (see Brame and Piquero 2003). For example, individuals who serve terms of confinement may lead highly mobile lives and become lost to survey staff over time. Fourth, our estimates are based on a national probability sample of young people, with incarceration identified prospectively. The average offender in this type of sample may be less active than the average offender in an inmate sample or an incoming cohort of offenders. Fifth, ours is a contemporary sample of incarcerated individuals who underwent confinement at the close of the 20th century (1998–2002). Because earlier bottom-up studies recruited offenders in the 1970s and 1980s, when U.S. incarceration was but a fraction of its current rate, we can anticipate a decline in crime frequency of offenders on the margin because of the diminishing returns phenomenon. Finally, because our counterfactual offending rate is based on self-reports of non-incarcerated individuals from a large community sample, all of the usual caveats associated with self-report criminal behavior apply.

In spite of these potential weaknesses, we view our method as a promising one for future applied research. The technique of propensity score matching could profitably be applied to a number of additional research problems related to justice system involvement. First, the effects of justice system involvement on future life outcomes such as offending (e.g., deterrence), employment, education, fertility, marriage, and welfare dependency can be assessed using this method. Second, the analysis can be expanded to look at the relative effects of different levels of justice system involvement: Long-term versus short-term incarceration, probation versus prison, arrest-only vs. conviction, etc. Additionally, the method could be applied to administrative data provided it included at least two levels of justice system sanctions and a sufficiently rich set of background characteristics to estimate propensity scores.

Prior studies often relied upon administrative samples to estimate the expected number of crimes or arrests averted through incapacitation. We use a population survey in which arrest, incarceration, and offending are self-reported. Thus, our incapacitation estimates refer to the typical experience of a 16–19 year old who becomes confined. The many methodological differences between this and prior studies may account for our lower estimates of the incapacitation effect. However, part of the difference may in fact stem from diminishing returns of the expanded use of incarceration. Had this study been conducted with a population sample from the 1970s, the prevalence of incarceration would likely be less than half that of the current sample, and the incarcerated individuals, on average, would be more serious offenders. Further research on earlier samples would help to resolve this issue.

Appendix: Standardized and Adjusted Bias for Each Type of Matching

Variable	Unmatched		Arrested		Convicted		Nearest neighbor		Kernel	
	16–17	18–19	16–17	18–19	16–17	18–19	16–17	18–19	16–17	18–19
Male[a]	60.2	83.4	26.1	42.2	27.9	37.7	4.2	−6.5	11.7	5.1
Hispanic[b]	5.7	−1.4	11.7	−1.3	8.2	−2.8	16.0	12.1	2.5	7.3
Black[a]	14.0	33.8	5.1	37.1	17.6	55.9	−14.7	−10.4	1.6	2.5
Nonwhite[b]	3.3	−12.9	7.4	−10.2	7.9	−8.9	13.1	−8.3	1.7	3.6
Rural, age 12	−4.4	−13.5	7.0	−6.7	32.5	−2.7	−5.2	−17.8	−3.2	−10.5
Urban, age 12	−7.9	0.4	−15.9	−4.8	−26.3	−1.1	−9.6	12.7	−4.3	2.7
Dwelling: house[a]	−23.3	−8.6	−21.5	1.8	−16.8	10.6	16.3	17.0	−2.0	11.6
Dwelling: apartment[a]	18.9	4.9	12.9	−1.0	14.1	−2.8	−6.8	−20.9	1.8	−11.4
Dwelling: other	9.9	6.8	15.9	−1.4	6.6	−13.1	−16.3	2.5	0.6	−2.3
Prop offenses, age 12	13.7	17.6	−2.5	9.1	4.8	12.6	−18.4	−2.7	−6.1	6.0
Theft, age 12	34.0	25.6	9.3	7.8	−4.5	−7.6	−12.9	14.0	6.6	16.9
Other property offenses, age 12	16.6	25.1	−10.1	14.4	−12.4	2.5	−19.9	23.0	−11.8	16.9
Attack others, age 12	27.7	12.2	27.1	8.2	12.1	5.7	16.7	−12.2	9.7	−4.3
Drug sales, age 12	29.1	13.7	5.6	−7.3	−23.7	−21.6	4.1	−14.1	2.3	−11.5
Delinquency variety, age 12[a]	56.2	49.4	6.0	4.9	−19.2	−19.5	−9.9	18.3	1.1	12.9
House burgled before age 12[b]	12.0	9.0	−3.7	−0.2	−3.4	−14.1	−2.5	1.9	−6.1	5.8
Saw someone shot before age 12[b]	36.6	44.2	15.5	31.6	−3.3	36.7	−2.5	−11.0	−6.6	−3.7
Was bullied before age 12[b]	6.2	15.1	−18.2	8.9	−12.5	8.8	−4.7	13.5	−18.2	9.6

continued

Variable	Unmatched		Arrested		Convicted		Nearest neighbor		Kernel	
	16–17	18–19	16–17	18–19	16–17	18–19	16–17	18–19	16–17	18–19
Threatened at school, w1[b]	25.8	27.0	−5.8	5.1	−20.0	1.7	−12.9	16.0	−14.1	8.5
Stolen from at school, w1[b]	13.3	15.9	1.2	−4.6	11.3	0.1	−19.2	21.9	−10.0	7.2
Fought at school, w1[a]	88.7	54.0	49.8	24.3	39.7	20.5	−12.9	21.0	3.0	7.2
School late twice in last year, w1[b]	51.4	37.8	7.7	13.4	−12.8	2.2	18.2	0.0	17.1	2.9
10+ school absences in year , w1[b]	38.1	47.4	21.7	30.6	−1.9	20.1	4.9	15.2	1.3	12.5
School attachment scale, w1[b]	−36.7	−38.9	−7.6	−18.4	5.9	−13.4	4.3	−2.3	−2.0	−9.9
Low grades (Cs or worse)[b]	62.1	50.9	19.6	22.2	9.1	20.4	−4.1	−13.6	8.7	0.4
High grades (Bs or better) [b]	−53.4	−66.3	−4.3	−28.0	10.4	−36.7	−2.3	5.2	0.3	−9.5
Low math knowledge, w1[a]	49.4	30.8	40.1	24.0	10.4	26.9	8.8	−20.2	9.9	−5.7
High math knowledge, w1	−47.7	−36.9	−16.2	−16.5	−15.6	−11.6	−8.9	8.6	−11.9	−4.5
Missing math knowledge test, w1	24.3	21.4	18.8	11.3	32.3	0.3	−16.0	9.9	−2.4	10.5
Low paragraph comprehension,w1[a]	46.1	44.1	24.7	43.3	25.5	52.9	2.2	0.0	6.3	9.4
High paragraph comprehension,w1	−45.9	−34.1	−9.2	−16.5	−10.2	−22.3	−4.0	5.6	−10.8	−3.8
Missing paragraph comp. test, w1	24.7	21.7	19.7	13.2	32.3	3.6	−11.4	10.0	−1.4	11.0
Low word knowledge, w1[a]	28.6	24.4	13.4	26.5	19.3	39.9	−11.6	−13.8	7.7	−2.7
High word knowledge, w1	−31.9	−34.4	−11.0	−21.7	−11.6	−20.8	−11.6	0.0	−6.4	−10.5
Missing word knowledge test, w1	24.9	21.8	19.8	13.2	32.3	3.6	−11.4	10.0	−1.4	11.0
Low arithmetic reasoning, w1[a]	44.4	28.8	33.5	31.1	33.6	42.3	2.3	−7.0	8.2	−2.6
High arithmetic reasoning, w1	−40.1	−38.3	−20.5	−24.2	−6.0	−18.4	−4.3	6.1	−8.1	−9.3
Missing arithmetic reas. test,w1	25.0	21.9	19.8	13.3	32.3	3.6	−11.4	10.0	−1.3	11.3
Missing any AFQT test, w1[a]	24.3	21.4	18.8	11.3	32.3	0.3	−16.0	9.9	−2.4	10.5
Mother 18 or below at first birth[b]	15.4	6.2	6.4	−3.1	0.2	−2.8	9.5	5.4	−4.5	4.6
No health insurance, w1[b]	2.0	29.4	−2.7	22.6	−8.9	28.2	−11.6	14.8	−18.7	−1.5
Attachment to mother, w1[b]	−22.1	−39.9	−4.0	−29.9	3.6	−13.3	−11.4	−10.3	−1.5	−11.4
Mother's support, w1[b]	−19.9	−19.8	−7.2	−11.3	0.5	−4.3	3.4	−22.8	−5.3	−9.2
One parent is dropout[b]	41.5	20.4	35.7	12.0	24.4	9.3	8.3	0.0	13.1	−2.2
One parent is college grad[b]	−26.3	−57.2	−16.9	−47.4	−11.5	−57.5	12.1	−16.3	3.6	−11.6
Below poverty level, w1	28.2	37.0	30.2	34.0	7.8	21.5	0.0	21.6	6.1	12.1
1–2 times poverty level, w1	5.4	22.3	−0.3	15.8	−5.3	16.1	−10.1	12.3	−7.8	14.2
2–3 times poverty level, w1	−2.0	−4.8	−12.7	−1.7	−3.2	−4.8	−10.7	−12.2	2.5	1.0
3–4 times poverty level, w1	−20.4	−12.3	−18.4	−2.7	−12.8	−2.9	7.0	9.7	−9.6	3.1
4+ times poverty level, w1	−20.3	−40.3	−7.9	−35.7	3.8	−28.6	5.8	−14.2	1.4	−13.1
Household income unreported, w1	−2.0	−17.6	−0.9	−20.3	4.9	−10.9	8.8	−20.1	3.9	−19.3
Parent received AFDC, w1	32.6	35.5	15.9	16.0	−6.4	20.4	−12.2	−6.0	−2.9	2.2
Parent received Medicaid, w1	29.7	21.6	20.3	5.8	18.9	7.6	−8.2	−9.3	−1.5	−0.6
Parent received SSI, w1	2.9	20.9	−15.7	9.8	12.5	−0.9	−32.1	7.5	−14.9	4.5
Parent received food aid, w1	17.8	34.5	−1.8	22.9	−24.8	16.0	−21.2	2.8	−9.8	10.0
Anti-social peer scale, w1[a]	49.4	30.7	7.7	6.6	−11.7	−9.4	4.0	2.5	1.9	0.5
Rural	−1.7	−2.4	13.3	0.0	31.5	−6.0	−6.7	−4.9	5.3	1.2
Urban	−2.6	5.3	−20.3	2.3	−34.3	15.4	2.1	4.8	−2.4	2.1
Northeast region[a]	−27.8	−0.9	−22.2	−5.3	−25.9	−10.1	−5.8	−7.4	−0.7	−6.7
North-central region	3.2	1.6	−0.3	4.6	0.2	2.2	−6.8	6.7	−3.2	7.1
West region	14.0	−16.9	10.3	−12.1	7.5	−20.4	13.1	−12.5	10.5	−2.4
Out of United States	−5.1	−5.5	0.0	0.0	0.0	0.0	0.0	0.0	−1.9	−2.1
Delinquency variety[a]	88.7	72.4	24.3	13.3	−22.7	−28.3	3.8	9.8	8.5	11.5
Delinquency frequency	46.4	38.6	32.8	22.6	13.8	−3.4	11.7	11.2	9.9	7.5

continued

Variable	Unmatched		Arrested		Convicted		Nearest neighbor		Kernel	
	16–17	18–19	16–17	18–19	16–17	18–19	16–17	18–19	16–17	18–19
Property offenses	39.4	23.6	30.8	14.9	26.9	−5.7	21.6	−9.7	15.2	−2.9
Theft	38.2	25.0	32.0	14.4	18.1	−3.5	5.9	7.9	4.2	2.6
Attacking others	45.7	33.8	15.0	26.7	−18.9	20.8	8.3	11.7	1.5	18.2
Drug sales	34.9	42.9	17.6	21.9	−15.6	−8.2	−4.6	23.1	4.3	11.7
Arrests[b]	65.2	59.1	0.0	−0.4	−80.8	−42.2	23.8	12.5	20.7	17.1
Ever arrested[a]	67.9	69.8	3.2	22.5	−99.5	−16.3	12.6	13.8	14.5	16.4
Charged[b]	63.5	60.8	−26.9	−27.2	−184.6	−145.3	13.4	8.6	20.3	15.8
Ever charged	66.9	69.0	−20.0	2.0	−187.0	−90.1	12.5	13.6	17.7	17.7
Court appearance	64.5	58.4	−15.5	−70.9	−246.3	−252.6	12.1	4.5	22.6	11.3
Ever appeared in court	66.1	67.7	−10.7	−45.1	−232.1	−193.5	14.6	13.5	21.9	14.9
Convicted[b]	46.8	47.1	−19.5	−44.8	−336.0	−322.9	7.4	0.0	13.7	9.7
Ever convicted	45.1	53.7	−20.8	−28.8	−332.3	−253.2	3.7	6.7	7.9	7.3
High school dropout[b]	27.1	46.8	8.8	26.1	19.6	19.0	−20.4	24.4	−17.4	12.1
Suspended[a]	108.3	62.7	29.6	16.0	16.3	6.0	−2.2	14.6	1.5	8.9
Lives with both biological parents[b]	−51.9	−62.4	−9.6	−38.0	−5.4	−28.5	−10.1	3.1	−17.8	0.2
Independent from parents[a]	55.1	46.2	42.9	31.4	26.5	27.4	10.7	22.8	3.2	8.9
Household size	14.1	−0.6	17.4	3.4	18.1	9.3	−0.6	−1.2	7.6	1.8
Sibling/peer in gang[b]	38.2	33.8	−2.7	11.4	−39.2	−4.9	6.6	21.3	1.3	18.0
Ever belonged to gang[a]	62.8	48.2	34.7	24.0	−3.2	6.1	13.1	21.0	5.8	15.5
Belonged to gang in last year[b]	42.1	37.1	18.9	25.7	−16.1	13.4	17.9	5.8	−0.7	13.7
Years sexually active[a]	88.4	82.8	61.7	45.5	22.7	43.4	−21.6	4.0	−9.2	−0.9
Drank alcohol[b]	35.8	24.7	−21.4	−15.7	−33.5	−40.2	11.6	0.0	4.9	8.2
Smoked[b]	65.8	67.9	−6.0	3.1	−21.0	−13.5	0.0	−7.4	9.4	−1.0
Used marijuana[a]	64.7	70.5	4.9	7.0	−42.1	−19.0	10.6	22.4	5.1	13.8
Used cocaine[b]	35.7	23.4	30.2	2.0	12.5	−18.5	15.7	12.7	18.7	−6.5
Days of last 30 drank alcohol	31.2	40.0	6.9	12.5	−5.9	−7.8	11.6	23.1	6.9	18.3
Alcoholic drinks/day, last 30 days	25.9	32.4	9.1	9.4	6.7	−6.1	4.7	24.8	4.7	16.5
Alcohol before/at school/work	28.0	28.0	13.4	19.2	20.6	1.2	1.5	9.9	5.3	7.4
Days of last 30 binged	27.6	37.2	4.4	11.4	1.5	−19.3	−0.9	14.0	0.5	9.5
Days of last 30 used cigarettes	63.4	66.5	20.6	20.1	−11.2	−16.1	6.8	−1.7	4.5	13.0
Cigarettes/day, 30 days[a]	63.2	54.3	31.4	21.3	9.0	−7.9	12.6	2.9	7.7	10.6
Days of last 30 used marijuana	41.4	63.4	19.6	34.0	−13.7	12.7	9.5	25.8	10.3	19.2
Marijuana before/at school/work	34.5	49.3	24.3	33.6	−11.6	28.9	7.2	17.5	18.4	23.0
Times used cocaine in last year	18.0	11.2	12.4	−6.5	3.4	−21.2	18.0	−0.3	10.1	−6.6
Cocaine before/at school/work	15.1	18.7	14.9	9.2	−14.9	14.8	5.8	11.1	2.4	5.5
Interior of house: nice[b]	−44.6	−35.4	−21.1	−22.3	−0.3	−18.6	11.7	−4.3	−5.4	0.2
Interior of house: fair[b]	20.4	9.2	7.2	3.0	5.3	−6.0	−4.1	9.2	4.9	1.5
Interior of house: poor[b]	29.0	32.0	9.8	14.7	−22.4	21.3	−20.4	4.1	−6.9	6.2
Exterior of house: nice[a]	−45.6	−40.7	−26.4	−33.6	−4.0	−27.1	−2.0	11.4	−6.9	4.0
Exterior of house: fair[b]	9.7	22.6	−1.1	20.0	−7.6	9.8	−2.0	−5.9	−2.3	−4.7
Exterior of house: poor[a]	38.8	25.3	26.3	13.9	4.0	19.8	−5.3	4.2	2.9	7.5
Missing victimization[b]	5.3	20.9	−37.6	−14.4	−22.1	−17.1	−13.2	−7.6	−6.0	−2.7
Missing school attachment[b]	10.9	19.7	9.4	14.4	16.5	19.3	−8.7	−5.0	10.4	0.3
Missing attachment to mother[b]	17.2	12.7	−2.4	−9.2	−7.4	−13.2	13.3	8.4	6.2	3.6
Missing mother's supportiveness[b]	30.4	22.6	22.4	10.6	27.5	−9.2	28.8	−5.6	20.7	−0.1

continued

Variable	Unmatched		Arrested		Convicted		Nearest neighbor		Kernel	
	16–17	18–19	16–17	18–19	16–17	18–19	16–17	18–19	16–17	18–19
Missing gang membership[b]	15.7	7.3	14.1	10.3	6.9	10.3	−9.5	−42.2	12.5	−42.5
Unbalanced (%)	69.40	71.80	31.50	37.30	33.30	33.60	7.20	16.40	4.50	1.80

All variables are measured in the year prior to potential treatment, unless otherwise noted

[a] Included in both 16–17 and 18–19 propensity score model

[b] Included in 18–19 propensity score model

References

Angrist JD, Imbens GW, Rubin DB (1996) Identification of causal effects using instrumental variables. J Am Stat Assoc 91:444–455

Ashenfelter O (1978) Estimating the effect of training programs on earnings. Rev Econ Stat 60:47–57

Avi-Itzhak B, Shinnar R (1973) Quantitative models in crime control. J Crim Justice 1:185–217

Blumstein A, Canela-Cacho JA, Cohen J (1993) Filtered sampling from populations with heterogeneous event frequencies. Manage Sci 39:886–899

Blumstein A, Cohen J, Roth JA, Visher CA (eds) (1986) Criminal careers and "career criminals", vol 1. National Academy Press, Washington, DC

Brame R, Piquero A (2003) Selective attrition and the age-crime relationship. J Quant Criminol 19:107–127

Campbell DT, Erlebacher A (1970) How regression artifacts in quasi experimental evaluations can mistakenly make compensatory education look bad. In: Hellmuth J (ed) The disadvantaged child, vol 3. Brumner/Mazel, New York, pp 185–210

Canela-Cacho JA, Blumstein A, Cohen J (1997) Relationship between the offending frequency (λ) of imprisoned and free offenders. Criminology 35:133–175

Chaiken JM, Chaiken MR (1982) Varieties of criminal behavior. Rand, Santa Monica, CA

Ehrlich I (1973) Participation in illegitimate activities: a theoretical and empirical investigation. J Polit Econ 81:521–565

Epanechnikov VA (1969) Non-parametric estimation of a multivariate probability density. Theory Prob Appl 14:153–158

Gibbs JP (1968) Crime, punishment, and deterrence. Southwestern Social Sci Quart 48:515–530

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford, CA

Greenwood P, Abrahamse A (1982). Selective incapacitation (Report R-2815-NIJ). Rand, Santa Monica, CA

Greenwood P, Turner S (1987). Selective incapacitation revisited (Report R-3397-NIJ). Rand, Santa Monica, CA

Harrison PM, Beck AJ (2006) Prison and jail inmates at midyear 2005 (NCJ 213133). U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC

Heckman JJ, Hotz VJ (1989) Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. J Am Stat Assoc 84:862–874

Heckman JJ, Smith JA (1999) The pre-programme earnings dip and the determinants of participation in a social programme: implications for simple programme evaluation strategies. Econ J 109:313–348

Hirschi T, Gottfredson M (1983) Age and the explanation of crime. Am J Sociol 89:552–584

Horney J, Marshall IH (1991) Measuring lambda through self-reports. Criminology 29:471–495

Kovandzic TV, Vieraitis LM (2006) The effect of county-level prison population growth on crime rates. Crim Public Policy 5:213–244

Leuven E, Sianesi B (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. http://ideas.repec.org/c/boc/bocode/s432001.html. Version 3.0.0

Levitt SD (1996) The effect of prison population size on crime rates: evidence from prison overcrowding legislation. Quart J Econ 111:319–351

Maltz MD, Pollock SM (1980) Artificial inflation of a delinquency rate by a selection artifact. Oper Res 28:547–559

Marvell TB, Moody CE (1994) Prison population growth and crime reduction. J Quant Crim 10:109–140

Nagin D (1978) General deterrence: a review of the empirical evidence. In: Blumstein A, Cohen J, Nagin D (eds) Deterrence and incapacitation: estimating the effects of criminal sanctions on crime rates. National Academy Press, Washington, DC, pp 95–139

Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. Criminology 31:327–362

Peterson MA, Braiker HB, Polich SM (1980) Doing crime: a survey of California prison inmates (Report R-220-DOJ). Rand, Santa Monica

Rolph JE, Chaiken JM (1987). Identifying high-rate serious criminals from official records (Report R-3433-NIJ). Rand, Santa Monica, CA

Rolph JE, Chaiken JM, Houchens RL (1981). Methods for estimating crime rates of individuals (Report R-2730-NIJ). Rand, Santa Monica, CA

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55

Rosenbaum PR, Rubin DB (1984) Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 79:516–524

Rosenbaum PR, Rubin DB (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. Am Stat 39:33–38

Shinnar S, Shinnar R (1975) The effects of the criminal justice system on the control of crime: a quantitative approach. Law Soc Rev 9:581–611

Smith JA, Todd P (2005) Does matching overcome LaLonde's critique of non-experimental methods? J Economet 125:305–353

Spelman W (1994) Criminal incapacitation. Plenum Press, New York

Spelman W (2000) What recent studies do (and don't) tell us about imprisonment and crime. In: Tonry M (ed) Crime and justice: a review of research, vol 27. University of Chicago Press, Chicago, pp 419–494

Tittle CR (1969) Crime rates and legal sanctions. Social Problems 16:409–423

Zimring FE, Hawkins G (1995) Incapacitation: penal confinement and the restraint of crime. Oxford University Press, New York