LABELING EFFECTS OF FIRST JUVENILE ARRESTS:

SECONDARY DEVIANCE AND SECONDARY SANCTIONING*

Akiva M. Liberman

Justice Policy Center, Urban Institute

David S. Kirk

University of Texas at Austin

KiDeuk Kim

Justice Policy Center, Urban Institute

Feb 2014 Manuscript. In press, *Criminology*

---

* This research was funded by the grant 2010-MU-FX-0613 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions expressed are those of the authors alone and do not necessarily represent the official positions of the U.S. Department of Justice. We are grateful to the Project on Human Development in Chicago Neighborhoods for providing the data necessary to undertake this study.

Labeling Effects of First Juvenile Arrests:

Secondary Deviance and Secondary Sanctioning

ABSTRACT

A growing literature suggests that juvenile arrests perpetuate offending and increase the likelihood of future arrests. The effect on subsequent arrests is generally regarded to be a product of the perpetuation of criminal offending. However, increased rearrest may also reflect differential law enforcement behavior. Using longitudinal data from the Project on Human Development in Chicago Neighborhoods (PHDCN) together with official arrest records, the current study estimates the effects of first arrests on both reoffending and rearrest. Propensity score methods were used to control differences between arrestees and non-arrestees and minimize selection bias. Among 1,249 PHDCN youth, 58 were first arrested during the study period; 43 of these arrestees were successfully matched to 126 control cases who were equivalent on a broad set of individual, family, peer, and neighborhood factors. We find that first arrests increased both the likelihood of subsequent offending and of subsequent arrest, through separate processes. The effects on rearrest are substantially larger and largely independent of the effects on reoffending, suggesting that labels trigger "secondary sanctioning" processes distinct from secondary deviance processes. Attempts to ameliorate deleterious labeling effects should include efforts to dampen their escalating punitive effects on societal responses.

INTRODUCTION

The 1980s and the early 1990s were characterized by an "epidemic" of youth violence in the United States, which peaked in 1993-1994 (Cook and Laub, 2002). Policy responses to the epidemic included a shift from the traditional rehabilitative goal of juvenile justice toward more retributive goals (e.g., Allen, 2000), under the mantra of "old enough to do the crime, old enough to do the time." The jurisdiction of the juvenile court was curtailed through lowered age of criminal responsibility, legislative exclusion of various age-charge combinations from juvenile court jurisdiction, and increased prosecutorial discretion to "direct file" cases in adult court (see Fagan and Zimring, 2000). The wisdom of retaining a separate juvenile court system was also debated (e.g., Ainsworth, 1995; Bishop, 2004; Butts and Mitchell, 2000; Dawson, 1990; Feld, 1998).

States have begun stepping away from the punitive philosophy of late 20th century juvenile justice (National Conference of State Legislatures, 2012; U. S. Department of Education, 2014), following the decline in youth crime and violence over the past two decades. Juvenile violent offending rates are now at historic lows, with the latest arrest data from the U.S. Office of Juvenile Justice and Delinquency Prevention (2012) down 55 percent from its mid-1990s peak. However, the decline in violent crime among juveniles outpaced the decline in arrests of violent juveniles through at least the early 2000s, so that the ratio of juvenile violent crime arrests to violent victimizations by juveniles increased from about 0.72 in 1980, to about 1.0 in the early 1990s, and to about 1.45 by 2003 (Snyder and Sickmund, 2006, p. 64). Despite massive declines in juvenile crime and violence over the past two decades, a convincing case can be made that U.S. society is still very much "governed through crime," with a youth control complex that criminalizes juveniles at an extraordinary level (Rios, 2011; Simon, 2007).

The current study examines the collateral consequences of this criminalization of youth, and revisits a question that has captivated and challenged criminologists for some time: What is the effect of arresting juveniles? Two theoretical perspectives provide opposing answers to this question. Deterrence theory predicts that arrests will have the specific deterrent effect of reduced offending (e.g., Smith and Gartin, 1989), while labeling perspectives predict that arrests will lead to increased offending and criminal sanctioning (e.g., Lemert, 1951). A third perspective, Gottfredson and Hirschi's (1990) self-control theory, argues that a lack of self-control explains any apparent relationship between system responses such as arrest and subsequent behavior, and that the relationship between juvenile arrest and reoffending is spurious. To date the empirical literature has revealed little support for specific-deterrence. The literature is largely split between null findings, in accord with self-control theory, and findings that seem to show that arresting juveniles is associated with more subsequent offending, in accord with labeling theory.

In view of the literature to date, a preliminary aim of this study is to test the replicability of the labeling effects previously reported. Through the use of propensity score methods combined with the necessary sensitivity analyses, this study aims to minimize the selection-bias threats to validity that are common in non-experimental studies. The study draws upon the broad data on youth, family, peer, and neighborhood characteristics collected as part of the Project on Human Development in Chicago Neighborhoods (PHDCN). The comprehensiveness of the PHDCN allows us to account for many confounding influences that distinguish arrestees from non-arrestees in estimating the relationship between juvenile arrest and future offending.

The primary aim of this paper is to then distinguish between two types of potential labeling effects: the effects of labels on delinquent behavior versus the effects of labeling on societal responses to the label, particularly via future sanctioning. This broadens our exploration of the effects of labels on not just deviant behavior but also effects on societal response to misbehavior.

Per these two aims, we ask the following research questions: Does the first arrest of a juvenile increase the likelihood of future offending? Does it increase the likelihood of subsequent arrest? Does juvenile arrest increase the likelihood of subsequent arrest even after accounting for any increases in offending? Put differently, does a first juvenile arrest lead to subsequent arrests even if the arrestee does not engage in more subsequent offending than a similar non-arrestee?

LABELING EFFECTS ON DELINQUENT BEHAVIOR AND ON SYSTEM RESPONSE

Labeling theory generally predicts that an "official" response to delinquency promotes future delinquency (e.g., Lemert, 1951). Labeling theory includes two different mechanisms by which a "label" can lead to increased deviancy (Paternoster and Iovanni, 1980). In one strand of labeling theory, the primary mechanism is that a delinquent label redirects a youth's self-conception or personal identity toward a deviant self-concept, which is then self-fulfilling (e.g., Matsueda, 1992). Edwin Lemert's (1951) version of labeling theory is emblematic of this process, particularly his depiction of the progression from "primary deviance" to "secondary deviance." Individuals come to internalize the deviant status stemming from societal reaction to their behavior, and deviants' come to organize their lives around this status (see also Becker, 1963; Schur, 1971). Labeled deviants may then associate with more deviant peers (Wiley Slocum, and Esbensen, 2013), withdraw from conventional pursuits (Bernburg, 2009), and ultimately engage in criminal offending at a higher rate than otherwise similar individuals who have not been labeled "deviant." With this higher rate of offending, stigmatized youth would presumably also have more frequent interaction with the criminal justice system than non-deviants.

Another mechanism in labeling theory focuses more on external processes involving social and societal responses to the label, including increased surveillance as well as reduced social opportunities and interactions (e.g., Klein, 1986; Link et al., 1989; Paternoster and Iovanni, 1989). Here, the mechanisms are not internal to the labeled individual, but rather the external social and societal responses, *per se*. In a parallel to Lemert's terms of primary versus secondary deviance, we conceptualize the labeling event—here an arrest—as a "primary sanction" and subsequent punitive societal responses resulting from the label as "secondary sanctioning." This terminology is intended to capture the idea that there may be two parallel processes operating in reaction to a deviant label, one internal and one external.

Representative of this version of labeling theory, Sampson and Laub's (1997) life-course theory of cumulative disadvantage emphasizes that once an individual is labeled a deviant, a variety of detachment processes are set in motion that promote the likelihood of further deviance. The stigma of a criminal record undermines social control processes, whether or not the labeled deviant internalizes the deviant status as in the Lemert framework. Sampson and Laub (1997, p. 147) note, "The theory specifically suggests a 'snowball' effect—that adolescent delinquency and its negative consequences (e.g., arrest, official labeling, incarceration) increasingly 'mortgage' one's future, especially later life chances molded by schooling and employment."

Several recent studies show evidence of such secondary sanctioning processes. For instance, Kirk and Sampson (2013) suggest that an arrest record officially marks a juvenile as a "criminal" and changes the way educational institutions treat the student. Students with criminal records are often pushed out of high school through exclusionary policies, and segregated into specialized programs for problem youth. The result of the primary sanction (arrest) and the secondary sanction (school exclusionary policies and practices) is an increased likelihood of high school dropout and diminished prospects for going to college (e.g., Bernburg and Krohn, 2003; Hirschfield, 2009; Kirk and Sampson, 2013; Sweeten, 2006), thereby leading to a higher likelihood of future criminality. Similarly, the stigma of a criminal record drastically influences how former offenders are treated by potential employers, and the denial of employment represents a form of secondary sanctioning (Laub and Sampson, 2003; Pager, 2003; Schwartz and Skolnick, 1962).

Moreover, if labeling effects operate though differential social or societal responses to those labeled as deviant, then a labeled individual may have more frequent interactions with the criminal justice system even if his or her criminal offending does not increase following an arrest (relative to otherwise similar "non-deviants" who avoided an arrest record). As Petrosino and colleagues put it (2010, p. 9), "The same actions that resulted in police turning a blind eye to misconduct may now result in an arrest." Such secondary sanctioning processes fit broadly under the realm of labeling theory, but offer slightly different predictions than classic versions of labeling which stress identity internalization, or even Sampson and Laub's (1997) version which stresses a decline in social controls. The essential difference is that the stigmatized deviant may not engage in crime at a higher rate following arrest relative to an otherwise similar individual who managed to avoid arrest, but the stigmatized deviant would still be rearrested and sanctioned more often because of the intensified gaze, or declining tolerance, of the criminal justice system.

PRIOR RESEARCH ON THE LABELING EFFECTS OF JUVENILE ARRESTS

Few studies of the effects of arrest, whether in experimental or observational studies, have simultaneously examined both secondary deviance (subsequent delinquency) and secondary sanctioning (subsequent justice-system responses). Most experimental studies have relied solely on administrative outcome data, and have generally taken official data (arrests) as an indicator of offending behavior *per se,* without distinguishing between effects on offending behavior (secondary deviance) versus effects on later system response (secondary sanctioning). Petrosino et al. (2010) recently conducted a meta-analysis of the effects of formal responses to juvenile delinquency, limited to studies with random assignment (or quasi-random assignment) of juveniles to either traditional processing versus release or some form of diversion. Overall, the meta-analysis found that formal sanctioning was associated with more reoffending, across self-report and official measures. Of the studies reviewed, 13 address the question of immediate interest here by comparing juveniles who received traditional processing – beginning with a formal arrest – to juveniles who were "released" or "counseled and released" without additional programming (see Petrosino, Turpin-Petrosino, and Guckenburg, 2010, Table 8.6).[1] All but one of these 13 experimental studies used official arrest measures.

One of the few experimental studies to measure both self-reported offending (SRO) and official arrests was conducted by Klein (1986). Youth identified by police were randomly assigned to be counseled and released, petitioned, or referred to one of two diversion conditions. Nine-months later, no effects were found on youths' SRO or their agreement with descriptions of themselves that "encapsulated" a delinquent label, but formally-petitioned youth were more likely to have been rearrested.

In contrast to the experimental studies, most longitudinal studies have relied upon SRO outcomes, and have not explored system responses. In a recent review, Huizinga and Henry (2008) identified 19 longitudinal studies of the effects of arrest with reasonable attempts to control for selection; most relied solely upon SRO. About half found no effect of arrest; the other half seemed to find labeling effects on SRO.

Longitudinal studies can allow confident establishment of temporal ordering. However, selection bias remains a persistent challenge to non-experimental labeling findings, even in longitudinal designs. In addition to lower self-control (Gottfredson and Hirschi, 1987), arrested youth typically differ from non-arrestees in many ways that predispose them to greater offending, including individual-level risk factors, as well as family, peer, and neighborhood risk. To control this selection bias, about half of the studies reviewed by Huizinga and Henry included predisposing factors to arrest as control variables in regression models, and about half used matching strategies.

One of the most extensive matched longitudinal studies was conducted by Huizinga, Schumann, Ehret, and Elliott (2003). Arrest during adolescence was examined in two longitudinal samples, from Denver and from Bremmen (Germany). Each arrestee was matched to a non-arrestee who was most similar on gender, age, minority status, annual delinquency since age 14 up to the age of arrest, history of prior arrest, and annual history of delinquent peer involvement.[2] With this matching, adolescent arrest showed little effect on subsequent SRO in either site, nor did sanctions beyond arrest show much effect on subsequent SRO in Denver,[3] but sanctions as an adolescent was related to increased unemployment as a young adult.

Propensity score methods have recently been used to better control selection bias, by allowing matching on many risk variables simultaneously (Rosenbaum, 2002). Rather than matching on a few select variables, one matches on a summary measure (i.e., the “propensity” of arrest) that is computed from many variables simultaneously. This addresses one limitation of traditional (exact) matching, where one typically can only match on a few variables simultaneously. Three recent longitudinal studies have applied propensity-score methods to the question of the labeling effect of arrest. Wiley and Esbensen (in press; also Wiley, Slocum, and

Esbensen, 2013) used student survey data from the second national evaluation (2006-2013) of the Gang Resistance Education And Training (GREAT) program. Controlling for seventeen pre-arrest covariates, they find that youth who report having been arrested subsequently self-report significantly more offending.

McAra and McVie (2007) used data from the Edinburgh Study of Youth Transitions and Crime to explore the effects of three stages of formal processing after police contact (charging by police, "referral to Reporter," and being brought to a hearing). Charging and referral did not affect SRO, but being brought to a hearing was associated with significantly more SRO. Thus, in a rare study where arrest and prosecution could be disentangled, the study found no effect of arrest *per se*, but a labeling effect of prosecution.

Labeling effects should theoretically be strongest for the first labeling event and each repeated labeling event should have a smaller marginal effect. The difference between having an arrest history or not should be larger than the difference between having three versus four prior arrests. Conversely, deterrence should also be strongest for the first arrest. Because most empirical studies fail to distinguish those newly labeled from those being labeled repeatedly, they may have inadvertently conducted weak tests of the marginal effects of additional arrests, (Paternoster and Iovanni, 1989). This may be one contributing factor to the many null findings in the literature.

One of the few studies to isolate first arrests is Morris and Piquero's (2013) analysis of National Youth Study data. Selection was controlled by using propensity score matching within groups with similar prior-offending trajectories (see Haviland et al., 2008). First arrests reported at wave 5 (1980) were found to increase SRO at wave 6 (1983). This effect was most substantial with chronic delinquents, and negligible for youth with very little prior offending.

In sum, the prior literature generally finds that arrest increases subsequent offending, as predicted by labeling theory, or finds no effect at all. Experimental studies generally show labeling effects on rearrest, but do not distinguish offending from sanctioning. Longitudinal studies show effects on SRO. Some recent studies reaffirm the basic labeling effect using propensity score methods to control for selection bias, but none of these recent studies have examined secondary sanctioning.

THE CURRENT STUDY

The prior literature does not clearly delineate whether the effect of arrest on rearrest is primarily the product of the indirect effect of arrest through subsequent offending, or whether the first juvenile arrest independently increases rearrest beyond any effect through future offending. This paper aims to distinguish labeling effects, if any, on reoffending (secondary deviance) from labeling effects on rearrest (secondary sanctioning).

To study the effects of juvenile first arrests on both subsequent SRO and rearrest, we use data from the PHDCN, linked to official arrest data from the Chicago Police Department (CPD) and the Illinois State Police (ISP). Using the PHDCN data, we construct propensity scores for matching arrestees and non-arrestees using a rich set of covariates, including measures of prior offending, temperament, family circumstances, demographics, education, peer influences, and neighborhood characteristics. In contrast to most prior studies, we restrict the present study to first arrests.

We test three hypotheses: (1) A first juvenile arrest has an independent positive effect on subsequent delinquency and criminal offending above and beyond the influence of individual, family, peer, and neighborhood, and school correlates. (2) Arrested adolescents are more likely to be arrested in the future than otherwise similar youth without arrest records. (3) The effect of first arrest on future arrests is independent of the effect of a first arrest on criminal offending. That is, adolescents with a previous arrest are more likely to be arrested in the future than comparable non-arrestees even if they engage in similar levels of future delinquency.

DATA AND METHODS

SAMPLE

Our sample comes from the PHDCN's longitudinal cohort study, which involved three waves of data collection from seven cohorts of youth, at three-year age intervals (i.e., ages 0, 3, 6, 9, 12, 15, and 18 at wave 1). The first wave of interviews was conducted in 1995 through 1997, and subsequent waves were separated by about 2.5 years, with the third interviews occurring approximately five years later (from January 2000 to January 2002).

For the longitudinal study, the PHDCN selected a sample of 80 neighborhood clusters, stratified by racial/ethnic composition (seven categories) and socioeconomic status (high, medium, and low), from a total of 343 neighborhood clusters in Chicago (Sampson, Raudenbush, and Earls, 1997). Within these 80 neighborhood clusters, a simple random sample of households yielded a total sample of 1,517 youth in the 12-year-old and 15-year-old cohorts. We specifically focus on these two cohorts because our study examines the effect of arrest as a juvenile.

For rearrest outcomes, our sample consists of 1,249 youth, 58 arrestees (the treatment group) and 1,191 non-arrestees, who completed the Wave 1 SRO questionnaire and consented to the official records search. This excludes 34 youth who failed to respond to the Wave 1 SRO questionnaire, as well as 234 youth who did not to consent to the official records search. Prior research comparing PHDCN respondents who did or did not consent to the records search found no systematic difference on a measure of self-reported arrest (Kirk, 2008).

Although Wave 3 participation was not necessary for our rearrest outcomes, for SRO outcomes Wave 3 attrition slightly reduced the sample to 53 arrestees and 951 non-arrestees. Wave 3 attrition was lower among arrestees (8.6%; 5 youth) than non-arrestees (20.2%; 240 youth). The sample of arrestees was distributed across 39 of the 80 neighborhood clusters; only one neighborhood cluster contained more than three arrestees.

DESIGN

Reliably establishing the temporal order of pre-treatment propensities, treatment, and outcomes is a key requirement for a methodologically strong quasi-experimental study of the causal effects of arrests. We capitalize on the three-wave structure of the PHDCN longitudinal study and dates in the official arrest data to insure that measures used to predict our treatment condition—first arrest—were indeed measured prior to treatment, and that the re-offending and re-arrest outcomes follow the treatment condition. Pretreatment characteristics used in propensity models were measured at Wave 1, the treatment was restricted to a window between the Waves 1 and 3, and our self-report outcome is measured at Wave 3. Arrest outcomes are contemporaneous with the Wave 3 SRO period.

TREATMENT VARIABLE

Administrative data on arrests from the CPD and the ISP were obtained and merged with the PHDCN data. These data span 1995 to 2001, and include both juvenile and adult arrest data throughout Illinois. Identifying information used in matching the data sets includes social security number, name, birth date, county and zip code, race and ethnicity, and gender. To construct our treatment variable, we determined whether each given PHDCN survey respondent was officially arrested as a juvenile (per the CPD and ISP data) for the first time sometime between their Wave 1 interview date and 15 months preceding the Wave 3 PHDCN interview. This allows a 3 month buffer against recall error before the beginning of the 12-month recall period used for the SRO outcomes in Wave 3 (i.e., at the Wave 3 interview, respondents were asked about offending at any point during the prior 12 month period). With approximately five years between the first and third wave of data collection, this arrest window was approximately 45 months (60 months – 15).

OUTCOME VARIABLES

One of our offending outcome variables is a general measure of SRO over the past year. Subjects were asked at the Wave 3 interview whether they had engaged in each of 22 behaviors during the preceding 12 months, and if so, how many times. As our general offending measure, we calculate the variety of offending across the 22 items, which counts the number of different types of criminal acts in which the person engaged in at least once. The 22 items consist of 6 violent offenses, 8 property offenses, 3 drug-selling offenses, 3 public-order offenses, and 2 status offenses.[4] In addition to the offending variety score, we also use as outcome variables separate measures of the prevalence of violent, property, and drug-sales offending.[5]

Official arrest records were used as the secondary sanctioning outcome, using the same arrest data from CPD and ISP described above. We constructed a binary prevalence measure indicating whether each individual was arrested at any point from 12 months prior to their Wave 3 interview date and the end of 2001, which was the last available extract of data on arrests from CPD and ISP.

PROPENSITY SCORE MATCHING

Most adolescents commit some delinquent acts (Porterfield, 1943; Short and Nye, 1957; 1958; Wallerstein and Wyle, 1947), but, as Tannenbaum observes (1938, p. 19), "Only some of the children are caught." Only a minority of delinquent acts are detected and fewer lead to arrests. Law enforcement officers exercise considerable discretion regarding whether and when to initiate a formal arrest. In their classic study, Black and Reiss (1970) found that only 15 percent of police contacts with juveniles resulted in an arrest. The arrest decision to a large extent, then, lies with the police and is based on a host of external factors beyond the criminal behavior of the arrested individual. In this sense, juvenile arrest has a random component, making it likely that for each arrested individual in the PHDCN sample, there are one or more equivalent non-arrestees, in terms of criminal offending and other pre-treatment covariates, who were fortunate enough to avoid arrest following the commission of their crime or crimes. We exploit this randomness in juvenile arrest via propensity score matching, to determine if arrest is causally related to subsequent offending and rearrest.

Using the PHDCN's extensive data on pre-treatment covariates, we used propensity score matching to identify comparison youth who were otherwise similar to treated (i.e., arrested) youth by modeling the probability of arrest. Propensity score methods allow the creation of balanced treatment and control groups who are equivalent on all measured covariates (Rosenbaum and Rubin, 1983). Given a set of covariates that account for the features associated with selection into the treatment condition, this approximates an experimental design (Morgan and Winship, 2007; Rosenbaum, 2002). Importantly, however, whether the set of covariates is sufficient to account for selection into treatment cannot be empirically determined. Instead, as with other regression or matching approaches, judgments of the sufficiency of the control variables must be assessed on a priori grounds based on theory and understanding of the treatment under consideration. In addition, sensitivity tests allow one to examine the sensitivity of the results to possible omitted variables. Use of sensitivity analyses is critical when employing propensity score methods, particularly in analyses where relatively few covariates are used to estimate the propensity score, but even in our analysis that draws upon a vast array of pretreatment covariates. We use Rosenbaum's (2002, 2010) bounding approach for this purpose.

Propensity scores were constructed from 79 Wave 1 covariates. SRO variables from the first wave of interviews, constructed identically to the SRO outcome variables from Wave 3 interviews, were included as pre-treatment covariates predicting the probability of arrest. We also included similarly constructed Wave 1 SRO variables concerning status and public order offending in constructing the propensity score. In addition, pre-treatment variables included drug use, temperament, household composition, parenting characteristics, socioeconomic status (SES), demographics, education, and peer influences. Neighborhood characteristics were also included, from the 1995 PHDCN Community Survey, U.S. Census data, and reported crime data (Sampson, Raudenbush, and Earls, 1997). To estimate the effect of a first juvenile arrest on future offending and arrest, we statistically match and then compare arrested and non-arrested sample members who are otherwise similar to one another in their frequency of criminal offending and all the pretreatment characteristics.

Before creating propensity scores, multiple imputation procedures were used to impute missing values for Wave 1 pre-treatment variables. Approximately 24 percent of the cases had at least one missing value. Multiple-chain imputation was implemented through Stata 12's MI procedure, to create five imputed data sets. We followed Hill's (2004, p. 13) multiple-imputation matching strategy and calculated a propensity score for each observation in each of the imputed data sets, using the mi estimate and mi predict commands in Stata 12. We then averaged the propensity scores for each respondent across the five imputed data sets.

Propensity score matching was done using nearest neighbor 3:1 matching, with replacement, with a caliper set at 0.02. That is, each arrested youth was matched with up to three otherwise similar non-arrested youth who had a propensity of arrest (i.e., a predicted probability of arrest) within .02 of the arrested youth. Matching was accomplished via Stata's psmatch2 routine. The resulting matched samples for arrest outcomes consisted of 43 arrestees and 126 nonarrestees; because we matched with replacement, the 126 control matches include 103 unique control cases. For SRO outcomes, the matched samples consisted of 38 arrestees and 111 non-arrestees; these control matches include 80 unique control cases.[6]

Linear models of the effect of first arrest on offending variety were estimated directly by psmatch2. With binary outcomes (the prevalence of offending as well as rearrest), psmatch2 was used to identify matches; then the matched samples were analyzed via logit models while accounting for matching. That is, each treatment observation and its corresponding control matches represent a cluster, and logit models were estimated while accounting for this clustering.[7]

RESULTS

PRE-TREATMENT DIFFERENCES

A key early question concerns the similarity or difference among youth who are and are not arrested. Arrested youth significantly differed from non-arrestees on 34 of 79 covariates examined. Table 1 shows differences in individual-level factors in the left panel, and Tables 2 and 3 display differences among family, peer, and neighborhood-level factors.

Prevalence of offending at Wave 1 among arrested youth was significantly greater across all offending categories (violence, 60%; property, 22%; drug-sales, 14%; public–order, 21%; status, 64%; mean offending variety = 4.28) than among control youth (violence, 15%; property, 9%; drug-sales, 2%; public–order, 11%; status, 12%; mean offending variety = 1.80). In addition, arrestees differed significantly on variables in other domains. Arrested youth were significantly more likely to smoke cigarettes, with a marginally significant difference in marijuana use as well. In terms of temperament, arrested youth had lower inhibitory control, more sensation seeking, and were quicker to make decisions (i.e., a form of impulsivity). As noted by Gottfredson and Hirschi (1990), these aspects of self-control are likely confounders related to both the likelihood of first arrest and future behavior. Accounting for these aspects of control and temperament are essential when attempting to estimate the unbiased effects of arrest.

Among demographic variables, arrested youth were significantly more likely to be male, older, and African-American, and were less likely to be Mexican or second-generation immigrants (compared to third or later generations). First-generation immigrants were also descriptively less likely to be arrested than third generation or later, although the difference was not statistically significant.

(Tables 1, 2 and 3 about here)

Table 2 shows differences in family and parent variables. Arrested youth were less likely to have married parents, had lower levels of parental supervision, and came from families with more parent-child and family conflict as well as less developmental stimulation and parental warmth.

Table 3 shows peer characteristics and neighborhood differences. Arrested youth reported significantly more peer deviance, more peer pressure, and less peer attachment. At the neighborhood level, arrested youth lived in neighborhoods with proportionally more African Americans residents, fewer Hispanics, and fewer foreign-born residents, as well as higher concentrations of poverty and higher violent crime rates. The neighborhoods had more social and physical disorder, and residents had significantly more cynicism toward the law.

PROPENSITY SCORES AND BALANCE

The maximum propensity score, expressed as the probability of being arrested, found for the non-arrestees was 0.738 while the lowest propensity for the treatment group was 0.003 (see Figure 1). Following recommendations of Ho, Imai, King and Stuart (2007) and Stuart (2010), we restricted our analyses to individuals with propensities in the ranges found in *both* groups, known as the region of common support. Thus, we excluded four arrestees with propensity scores greater that any controls (i.e., greater than 0.738). We do not extrapolate our results to individuals with the highest propensities to be arrested, leading us to estimate only the "Average effect of the Treatment on the Treated" (ATT; the effect of treatment for those subjects who actually received the treatment). These effects may not generalize to individuals with very high probabilities of arrest.

Once matched to arrestees, the resulting sample of non-arrestees did not differ from the arrestees on any of the covariates, as shown in the right half of Tables 1-3. The post-match t-statistics and corresponding p-values on the right side of Tables 1-3 reveal that among the 79 covariates used to estimate the propensity of arrest, none showed a significant difference between the treated and controls in our final matched sample. In addition, matching on propensity score reduced absolute bias across all covariates by 63%, from a mean of 30.3 down to 11.5 (median bias reduced from 25.8 to 11.1).[8] [9]

(Figure 1 about here)

EFFECT OF ARREST ON OFFENDING AND RE-ARREST

Having established the effectiveness of our propensity score matching to produce equivalent samples of arrestees and non-arrestees, we turn now to the results of the effect of first juvenile arrest on self-reported reoffending from the third wave of PHDCN subject interviews.

As expected, in the absence of matching, the prevalence of offending at Wave 3 among arrested youth was considerably greater (violence, 57%; property, 30%; drug-sales 17%) than among control youth (violence, 23%; property, 19%; drug-sales, 6%); offending variety too, was considerably greater for arrestees (2.08 vs. 0.90). And those already arrested were much more likely to be arrested later (55% vs. 9%). These differences, all highly significant, reflect both the preexisting differences in propensity between those who had and had not been arrested (i.e., selection effects) as well as any effects of being arrested.

The key question is whether these differences persist once the selection effects are reduced through propensity score matching. We find that arrestees continued to report significantly more offending variety at Wave 3 than matched non-arrestees (2.03 vs. 1.04; SE=0.40; t=2.44, p=.016). In addition, as shown in Figure 2, arrestees were more likely to report committing violent, property, and drug distribution offending than matched non-arrestees. These prevalence differences were significant in logit models for violent offending (OR = 3.23; SE=1.45; z = 2.61, p = .009) and marginally significant for property offending (OR = 2.17; SE=0.98; z = 1.72, p = .086). In addition, as shown in the figure, arrestees were much more likely to be rearrested than equivalent non-arrestees (OR = 5.20; SE=1.85; z = 4.63, p < .001).

(Figure 2 about here)

SENSITIVITY AND ROBUSTNESS

Sensitivity. We explored the sensitivity of our results to possible unobserved variables. The rich set of covariates used in our propensity score analyses suggests that our matched results control for pre-treatment differences between the treatment and control groups. Nonetheless, the potential for hidden biases in our estimation of the effect of arrest remains. We used Rosenbaum’s (2002) bounding strategy, which explores how large the bias of an omitted variable would need to be to substantively affect our results (see Appendix A for methodological details). Γ in Table 4 refers to the assumed increase in the odds of treatment (first arrest) due to hypothetical unobservable factors.

We begin with the effect on future arrest. At $\Gamma = 1$, we assume there are no hidden biases, and would conclude that arrest has a significant positive effect on future arrest ($Q+ = 4.320$, $p < .001$). Hypothetical unobserved variables that would bias the results in the direction of the observed effect are explored with values of Γ larger than 1. At $\Gamma = 1.25$, we assume there is an unobserved variable that increases the odds of being arrested (receiving the treatment in the current study) by an additional 25 percent after accounting for the propensity score. Under this scenario, we still find a significant positive effect of arrest on future arrest ($Q+ = 3.720$, $p < .001$). At $\Gamma = 2.00$, we assume an unobserved variable that doubles the odds of being arrested, and we still find a significant positive effect of arrest on future arrest ($Q+ = 2.496$, $p < .01$). To render the effect on future arrest insignificant would require a Γ value of over 2.75. As an example, increasing the probability of arrest from .50 (odds=1.0) to .7333 (odds=2.75) would produce a Γ of 2.75.

For the effect on future violent offending, we find that at $\Gamma = 1.25$, the effect persists ($Q+ = 1.992$, $p = .023$). To render the treatment effect of arrest on violent offending no longer significant, at $p < .05$, would require a Γ value of nearly 1.5. As an example, increasing the probability of arrest from .50 (odds=1.0) to .60 (odds=1.5) would produce a Γ of 1.5.

As a comparison, we find that Wave 1 violent offending increases the odds of first arrest by an additional 20 percent after controlling for a propensity score that excludes this factor. Thus, in order to spuriously produce the effect on future violent offending, an unobserved factor would need to be related to first arrests somewhat more strongly than is prior violent offending. To spuriously produce the effect on future arrest, an unobserved factor would need to be related to first arrests much more strongly than is prior violent offending. Given that we already control for a full range of offending behavior and substance use in developing the propensity score, it is challenging to conceive of an omitted factor that would yield such a sizable increase in the probability of arrest, particularly for the effect of arrest on future arrest.

Robustness. We also explored the robustness of the effects found in the matching approach already described (3:1 nearest neighbor matching, with caliper = 0.02), through four other propensity-score-matching specifications. We widened the caliper for 3:1 matching to 0.04, used 1:1 matching, and used kernel matching with bandwidths set at 0.06 or 0.10. The alternative matching approaches and their results are presented in Appendix A.

The matching approach already described, which we term our “primary” specification, was the most efficient in removing bias.. This primary specification was more stringent than most of the other specifications on matching, so that 43 of the 58 arrestees were matched (a.k.a., "on support"), and 38 of the 53 arrestees with SRO data. The other specifications were somewhat less efficient in reducing bias, but were able to match more treated cases to controls.

The estimates of the effect on future arrest were highly significant across all specifications, and effect sizes (ORs) were larger under the alternative specifications than with our primary specification. However, for the effects on future offending, the picture was more mixed. The effects on violent offending were largest under our primary specification, and were reasonably robust, although the ORs and significance were somewhat reduced under other specifications. The results on property offending and on offending variety score were less robust, and effects were considerably diminished and no longer significant using kernel matching rather than nearest neighbor matching.

In sum, we find that the effects on future arrests effects are robust and are not sensitive to omitted variable bias. The effects on future offending, however, were both less robust and somewhat more sensitive. The largest offending effect, on violent offending, is reasonably robust and reasonably insensitive to omitted variable bias.

(Table 4 about here)

THE RELATIONSHIP BETWEEN SECONDARY DEVIANCE AND SECONDARY SANCTIONING

To this point, we have evidence in support of our first two hypotheses: arrest has an effect on offending, particularly violence, and also on rearrest. We now consider whether the effect of arrest on rearrest is independent of its effect on offending. Put differently, are individuals with an arrest record more likely to be arrested in the future even if they engage in comparable rates of offending as non-arrestees? Or, are prior arrestees more likely to be rearrested in the future primarily because of greater offending?

That the difference in rearrest reported in Figure 2 is considerably larger than the differences in SRO suggests that the two effects are distinct. To explore these questions more formally, we conducted an analysis of rearrest while controlling for the SRO at Wave 3. For this purpose we use our most robust effect of offending, violent offending, as a mediating variable. We weight cases by the inverse of their propensity score, and estimate the effect of arrest on subsequent arrest, net of criminal offending, in a logistic regression model (see Appendix A for methodological details).

Results are shown in Table 5. Model 1 replicates our earlier propensity score matching results with propensity weighting, and confirms that first arrests significantly increase the likelihood of later arrest (OR = 4.97).[10] Model 2 then explores the extent to which this effect is accounted for by increased SRO.[11]

Consistent with our third hypothesis, the effect of first arrest on future arrest is essentially independent of the effect on reoffending. That is, offending at Wave 3 is not predictive of rearrest (with a nonsignificant negative coefficient), once common influences are controlled through propensity scores. The clear implication is that the secondary sanctioning effect of increased rearrest is distinct from the secondary deviance effect on reoffending.

(Table 5 about here)

Finally, Model 3 examines whether the system response effect (i.e., the significant effect of arrest on subsequent arrest) is concentrated among the more active offenders, by interacting arrest with the level of offending. Focusing on the interaction term, we find no evidence that arrest is any more or less consequential for high rate versus low rate offenders.[12]

DISCUSSION

This study explored the effects of arresting juveniles on subsequent offending and on rearrest, and tested two types of labeling effects. Supporting our first hypothesis, we found that arrest led to a greater likelihood of offending, consistent with labeling theory. Supporting our second hypothesis, first arrests increased the likelihood of rearrest, and this effect was considerably larger than the effects on subsequent offending. Supporting our third hypothesis, we found that the increased likelihood of subsequent arrest was not due to the increase in offending. Rather, a first juvenile arrest appears to increase subsequent law enforcement responses to those youth compared to other youth who are offending at a comparable level but managed to evade a first arrest. This could result from increased scrutiny of the individual's future behavior, by police as well as other actors such as teachers and school staff, as well as through reduced tolerance by police and other actors of an arrestee's future transgressions.

These findings extend prior labeling research in several important ways. The increases in SRO following arrest confirm findings from studies with other longitudinal data. Using propensity score methods with the PHDCN's extensive set of covariates from the individual, family, peer, and neighborhood domains provides grounds for believing that we have substantially reduced the threat to validity from selection bias that is a concern in many labeling studies. Varied specifications and sensitivity tests confirm that our results are reasonably robust, especially the effect on violent offending, and relatively insensitive to bias from unobservables. These findings are also consistent with two other recent propensity score studies of SRO outcomes, Morris and Piquero's (2013) study of arrests circa 1980, and Wiley and Esbensen's (in press) study of arrests in the 2000s, both of which support the conclusion that arrests lead to secondary deviance.

The present study also confirms earlier findings, mostly from experimental studies, that an arrest tends to generate more subsequent arrests. This finding is robust across model specifications and quite insensitive to the possibility of omitted variable bias. Perhaps our most important finding concerns the relationship between the effect on SRO and the effect on rearrest. It is commonly assumed that rearrest is largely a product of the perpetuation of offending associated with secondary deviance. Our findings cast doubt on this common interpretation. Instead, we find a considerably larger effect on arrest than on SRO, consistent with Klein (1986).[13] Moreover, we find that the arrest effect is not diminished after accounting for the potential mediating effect of Wave 3 SRO, leading us to conclude that the effects of secondary deviance and secondary sanctioning are essentially independent.

One limitation of the present study is that it excludes the highest rate offenders, for whom no matched non-arrestees were found with a comparable propensity to be arrested. Hence, we have no empirical basis to estimate how such extremely high-propensity youth would fare if not arrested, because all such youth in the PHDCN were arrested. Possibly, the highest propensity youth are already so firmly on a deviant trajectory that a formal arrest has little effect either on their behavior or on societal responses. Or perhaps cumulative disadvantage makes first arrests especially likely to result in secondary deviance and/or secondary sanctioning for such high propensity youth.

IMPLICATIONS

Our results suggest that the large labeling effect found on rearrest truly reflects secondary sanctioning—that is, differential societal response to a youth with an “arrestee” or “delinquent” label—and that this societal response is not mediated by differential offending behavior of the juvenile. This process of secondary sanctioning, in which initial arrests beget further arrests, suggests a cascading effect of deepening involvement in the justice system.

How does this secondary sanctioning come about? Does it reflect only differential behavior of law enforcement or might some differential youth behavior also be implicated? One limitation from applying data collected for broader purposes to these labelling questions is that we have limited information about the arrests, their particular circumstances or what offending behavior may have prompted those arrests, let alone how the arrestee interacted with law enforcement. Thus, although the secondary sanctioning effects were not mediated by differential offending, we cannot confidently assume that no other behavioral differences between youth with and without prior arrest experiences played a role in their differential subsequent arrests. Nonetheless, the major expected behavioral contributor of youth would seem to be offending differences. The absence of any indication that offending behavior mediates the increase in arrest suggests, therefore, that secondary sanctioning effects are due in considerable part to differential societal responses. Whether this reflects police actively scrutinizing and/or monitoring "the usual suspects," being more likely to take formal action rather than issue a warning to youth with a prior arrest, or some other changed response remains a question for future research.

Labelling effects on youth behavior and on police behavior may also be interdependent. Secondary sanctioning processes may be partly co-produced in the interaction between two actors whose expectations are conditioned by the earlier arrest, with police having higher expectations that labeled youth will offend, while youth have higher expectations that police will carry out arrests. Long-standing experimental work finds that experimentally induced expectations of students' academic performance influence teachers' and students' interactions in ways that are expectancy-confirming (Jussim, 2005; Rosenthal and Jacobson, 1968), especially when teachers do not yet have much direct experience with the student (Raudenbush, 1984). Similar expectancy processes may operate in how labels affect the interactions between labelled offenders and justice system actors.

The present study thus suggests that understanding labeling processes will require future work on mechanisms of secondary sanctioning processes as well as on secondary deviance. This will require broadening the labeling perspective to include studying the behavioral response of societal actors to labels.

The policy implications of labeling findings are twofold. The most obvious implication for curtailing the destructive effects of labeling is to restrict formal law enforcement responses to serious delinquency and to resist the temptation to criminalize minor misbehavior, such as school discipline problems (e.g., Kupchik, 2010). While few U.S. policy-makers have been willing to go as far as Schur (1973) in considering "radical non-intervention" by law enforcement, there are some promising signs in the school domain. Following the lead of several school districts, notably in Broward County, Florida (see Stucki, 2013), the U.S. Department of Education (2014) has recently issued a set of guiding principles with respect to school discipline that marks a fundamental shift away from the tough-on-crime school policies that led to the criminalization of so many minor school infractions over the past two decades.

The second type of policy implication concerns how to decrease the detrimental effects of labeling. This typically focuses on how to ameliorate secondary deviance effects (e.g., Burnberg, Krohn, and Rivera, 2006; Wiley, Slocum, and Esbensen, 2013). The current study, however, suggests that in the relatively short term, the effects of arrest through secondary deviance may be dwarfed by the detrimental effects that operate though secondary sanctioning. Ameliorative policy efforts that address the secondary societal responses may be as important as those that try to ease the deviance-amplifying effect of the primary sanction. In the law enforcement realm, these would involve efforts to prevent the compounding effect of increasingly punitive law enforcement responses to equivalent misbehavior, especially minor misbehavior. These results also highlight the importance of policies and practices to maintain the confidentiality of juvenile records, and expunge and remove such records after an appropriate period of redemption (Blumstein and Nakamura, 2009).

One way for an offender to lessen the risk of secondary sanctioning is to simply move residences or change schools. There may be some benefit to putting physical distance between an arrestee seeking to reform his or her behavior and the increasingly watchful eye of the authorities. Recent research on prisoner reentry has found that residential relocation can potentially provide a turning point for ex-prisoners by helping sever ties to former peers and neighborhoods, thereby lessening some of the risk factors which propel individuals toward crime (Kirk, 2009; 2012). Residential change may also lower the risk of rearrest by separating individuals from the watchful gaze of local police who might have a decreasing level of tolerance for a given person’s transgressions. In accord with this line of reasoning, Keels (2008) finds that male youths from families participating in the Gautreaux housing mobility program in Chicago who moved to the suburbs were significantly less likely to be arrested for drug, theft, and violent offenses than male youths who moved internally within Chicago. A move to the suburbs means moving outside of the jurisdiction and surveillance of the Chicago Police Department, thereby providing more of an opportunity for a fresh start.

In conclusion, among otherwise equivalent youth with similar levels of criminal offending, those youth unlucky enough to become ensnared by the criminal justice system face a daunting task of steering clear from future interaction with the system. Not only is the likelihood of future offending increased for a host of reasons, but the likelihood of future sanctioning increases even if criminal behavior does not escalate relative to non-arrested counterparts. Therefore, policy solutions to the detrimental consequences of a delinquent label must address not only ways to reduce secondary deviance, but also ways to reduce secondary sanctions. Of course two ways to avoid the necessity of countering the consequences of criminal stigma are to reduce primary deviance and primary sanctioning. Fortunately, the federal government has recently recognized the far reaching consequences of a criminal stigma, and has taken recent steps to reduce the criminalization of youth, particularly misbehavior in schools (U.S. Department of Education, 2014). The U.S. Department of Education’s new guiding principles are a refreshing alternative to the culture of control that has characterized the U.S criminal justice and education systems for too long.

FOOTNOTES

[1] The extent of formal processing varies among these studies. "Because the system processing condition is usually the control group in the experiments, it often is not described further" (Petrosino, Turpin-Petrosino, and Guckenburg, 2010:13), and was described just as "processing" in nearly two thirds of the studies reviewed (p. 22). Among studies in which the control group was "released" or "counseled and released," the commonality among the traditional processing conditions seems to be a formal arrest. This is similar to studies of "arrest" *per se*, in which system penetration following arrest is generally unknown.

[2] Neighborhood type was also used as a matching variable in the Denver sample.

[3] In Bremmen, too few youth were sanctioned to support analyses.

[4] *Violent Offenses*: carried a hidden weapon, hit someone not lived with, attacked someone with a weapon, used force to rob, threw objects at people, was in a gang fight. *Property Offenses*: damaged property, set a fire, broke into a building to steal, stole from a store, stole from household member, snatched a purse, stole from a car, bought/sold stolen goods. *Drug-Selling Offenses*: sold marijuana, sold cocaine/crack, sold heroin. *Public-Order Offenses*: caused trouble in public, paid for sex, got a driving ticket. *Status Offenses*: ran away from home overnight, drove without a license.

[5] We also explored the frequency of offending in these three items. Results from analysis of offending frequency were substantively similar to the prevalence results.

[6] Reestimating the rearrest model using the smaller SRO sample generated equivalent results to those presented below (available upon request).

[7] Stata code for the logit models is as follows: *logit outcome_var treatment_var [pweight=controlweight], vce(cluster clusterid)*. Each treated case and its matched control cases have a common, unique cluster identification (*clusterid*); *vce(cluster)* accounts for this clustering. Control cases are down-weighted if more than 1 case match a given treated case; *controlweight* =1 for 1 match, .5 for 2, and .333 for 3 matches. Treated cases' weight = 1.

[8] Bias represents the mean differences across groups as a percentage of the square root of the average of the sample variances: $100 * (\bar{x}_T - \bar{x}_C)/(s_T^2 + s_C^2)^{1/2}$, where $\bar{x}_T$ and $\bar{x}_C$ are the sample means in the treated group and the control group respectively, and s_T^2 and s_C^2 are the respective sample variances (Rosenbaum and Rubin, 1985). Absolute bias is unsigned, and facilitates comparison across variables.

[9] Tables 1-3 matching results are based on the sample of youth with Wave 3 SRO data on violent offending. With the slightly larger sample of youth with official (re)arrest data, 3:1 matching reduced mean absolute bias by 66%, from 29.1 to 6.9, and no significant covariate differences remained.

[10] This OR is slightly different than presented earlier because of differences in estimation method. Propensity score weighting utilizes the full sample under common support in estimating the effect of arrest, including control and treated cases not used in the matched samples.

[11] Our regression estimation of the indirect path to rearrest via increased offending does not control for possible alternative indirect pathways. Its identification is therefore weaker than for our PSM estimation of the primary effects. However, because this mediating path is the critical pathway suggested by theories of secondary deviance, it warrants particular scrutiny. By

failing to control alternative pathways, we likely overestimate the magnitude of this pathway. In which case, this produces a conservative estimate of the residual effect that we take as evidence for a secondary sanctioning path.

[12] Similar results were founds in equivalent propensity-weighted analyses using the variety score as the control variable, or using offending from Wave 1 (rather than Wave 3).

[13] Although Klein found no effect on secondary deviance, we find such an effect on offending, perhaps because we restrict our study to first arrests, which should theoretically show the largest effects on secondary deviance.

REFERENCES

Ainsworth, Janet E. 1995. Youth justice in a unified court: Response to critics of juvenile court abolition. Boston College Law Review 36:927–51.

Allen, Francis A. 2000. Foreword. In *The Changing Borders of Juvenile Justice: Transfer of Adolescents to the Criminal Court*, eds. Jeffrey Fagan and Franklin Zimring Chicago, IL: University of Chicago Press.

Becker, Howard S. 1963. *Outsiders: Studies in the Sociology of Deviance*. New York: Free Press.

Becker, Sascha O., and Marco Caliendo. 2007. Sensitivity analysis for average treatment effects. Stata Journal 7:71–83.

Bernburg, John Gunnar. 2009. Labeling theory. In *Handbook of Crime and Deviance*, eds. Marvin D. Krohn, Alan J. Lizotte, and Gina Penly Hall. New York: Springer.

Bernburg, Jon Gunnar, and Marvin D. Krohn. 2003. Labeling, life chances, and adult crime: The direct and indirect effects of official intervention in adolescence on crime in early adulthood. Criminology 41:1287–318.

Bernburg, Jon Gunnar, Marvin D. Krohn, and Craig J. Rivera. 2006. Official labeling, criminal embeddedness, and subsequent delinquency: A longitudinal test of labeling theory. Journal of Research in Crime and Delinquency 43:67–88.

Bishop, Donna M. 2004. Injustice and irrationality in contemporary youth policy. Criminology & Public Policy 3:633–44.

Black, Donald J., and Albert J. Reiss, Jr. 1970. Police Control of Juveniles. American Sociological Review 35:63–77.

Blumstein, Alfred, and Kiminori Nakamura. 2009. Redemption in the presence of widespread criminal background checks. Criminology 47:327–359.

Butts, Jeffrey A., and Ojmarrh Mitchell. 2000. Brick by brick: Dismantling the border between juvenile and adult justice. In *Boundary Changes in Criminal Justice Organizations*, ed. Charles M. Friel. Washington, DC: National Institute of Justice.

Caliendo, Macro, and Sabine Kopeinig. 2008. Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys 22:31–72.

Cook, Philip J., and John H. Laub. 2002. After the epidemic: Recent trends in youth violence in the United States. Crime and Justice 29:1–37.

Dawson, Robert O. 1990. The future of juvenile justice: Is it time to abolish the system? Journal of Criminal Law and Criminology 81:136–55.

DiPrete, Thomas A., and Markus Gangl. 2004. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. Sociological Methodology 34:271–310.

Fagan, Jeffrey, and Franklin Zimring (Eds.). 2000. *The Changing Borders of Juvenile Justice: Transfer of Adolescents to the Criminal Court*. Chicago, IL: University of Chicago Press.

Feld, Barry C. 1998. Abolish the juvenile court: Youthfulness, criminal responsibility, and sentencing policy. Journal of Criminal Law and Criminology 88:68–136.

Gottfredson, Michael R., and Travis Hirschi. 1990. *A General Theory of Crime*. Stanford, CA: Stanford University Press.

Haviland, Amelia, Daniel S. Nagin, Paul R. Rosenbaum, and Richard E. Tremblay. 2008. Combining group-based trajectory modeling and propensity score matching for causal inferences in nonexperimental longitudinal data. Developmental Psychology 44:422–436.

Hill, Jennifer. 2004. Reducing Bias in Treatment Effect Estimation in Observational Studies

Suffering from Missing Data. Columbia University Institute for Social and Economic Research and Policy (ISERP) working paper 04-01. Available at: http://academiccommons.columbia.edu/catalog/ac:129151.

Hirano, Keisuke, and Guido W. Imbens. 2001. Estimating of causal effects using propensity score weighting: An application to data on right heart catheterization. Health Services and Outcomes Research Methodology 2:259–278.

Hirano, Keisuke, Guido Imbens, and Geert Ridder. 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71:1161–89.

Hirschfield, Paul J. 2009. Another way out: The impact of juvenile arrests on high school dropout. Sociology of Education 82:368–93.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis 15:199–236.

Huizinga, David, and Kim L. Henry. 2008. The effect of arrest and justice system sanctions on subsequent behavior: Findings from longitudinal and other studies. In *The Long View of Crime*. ed. A.M. Liberman. New York, NY: Springer Science+Business Media, LLC.

Huizinga, David, Karl Schumann, Beate Ehret, and Amanda Elliott. 2003. The effect of juvenile justice system processing on subsequent delinquent and criminal behavior: A cross-national study. Final report to The National Institute of Justice, Grant Number 199IJCX0037.

Jussim, Lee, and Kent D. Harber. 2005. Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. Personality and Social Psychology Review 9: 131–155.

Keels, Micere. 2008. Second-generation effects of Chicago's Gautreaux residential mobility program on children's participation in crime. Journal of Research on Adolescence 18:305–52.

Kirk, David S. 2008. The neighborhood context of racial and ethnic disparities in arrest. Demography 45:55–77.

Kirk, David S. 2009. A natural experiment on residential change and recidivism: Lessons from Hurricane Katrina. American Sociological Review 74:484–505.

Kirk, David S. 2012. Residential change as a turning point in the life course of crime: Desistance or temporary cessation? Criminology 50:329–58.

Kirk, David S., and Robert J. Sampson. 2013. Juvenile arrest and collateral educational damage in the transition to adulthood. Sociology of Education 86:36–62.

Klein, Malcolm W. 1986. Labeling theory and delinquency policy: An experimental test. Criminal Justice and Behavior 13:47–79.

Kupchik, Aaron. 2010. *Homeroom Security: School Discipline in an Age of Fear*. New York: New York University.

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

Lemert, Edwin M. 1951. *Social Pathology: A Systematic Approach to the Theory of Sociopathic Behavior*. New York: McGraw-Hill.

Link, Bruce, Francis Cullen, Elmer Struening, Patrick Shrout, and Bruce Dohrenwend. 1989. A modified labeling approach to mental disorders: An empirical assessment. American Sociological Review 54:400–23.

Lopes, Giza, Marvin D. Krohn, Alan J. Lizotte, Nicole M. Schmidt, Bob E. Vasquez, and Jon

Gunnar Bernberg. 2012. Labeling and cumulative disadvantage: The impact of formal police intervention on life chance and crime during emerging adulthood. Crime and Delinquency 58:456–88.

Mantel, Nathan, and William Haenszel. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 22:719–48.

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

McAra, Lesley, and Susan McVie. 2007. Youth justice? The impact of system contact on patterns of desistance. European Journal of Criminology 4:315–45.

Morgan, Stephen L., and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Morris, Robert G., and Alex R. Piquero. 2013. For whom do sanctions deter and label? Justice Quarterly 30:837–68.

National Conference of State Legislatures. 2012. *Trends in Juvenile Justice State Legislation: 2001-2011*. Washington, DC: National Conference of State Legislatures

Pager, Devah. 2003. The mark of a criminal record. American Journal of Sociology 108:937–75.

Paternoster, Raymond, Robert Brame, Paul Mazerolle, and Alex Piquero. 1998. Using the correct statistical rest for the equality of regression coefficients. Criminology 36:859–866.

Paternoster, Raymond, and Leeann Iovanni. 1989. The labeling perspective and delinquency: An elaboration of the theory and assessment of the evidence. Justice Quarterly 6:359–94.

Petrosino, A., C. Turpin-Petrosino, and S. Guckenburg. 2010. Formal system processing of juveniles: Effects on delinquency. Campbell Systematic Reviews 2010:1.

Porterfield, Austin L. 1943. Delinquency and outcome in court and college. American Journal of

Sociology 49:199–208.

Raudenbush, Stephen W. 1984. Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. Journal of Educational Psychology 76: 85–97.

Rios, Victor. 2011. *Punished: Policing the Lives of Black and Latino Boys*. New York: NYU Press.

Rosenbaum, Paul R. 2002. *Observational Studies.* New York: Springer.

Rosenbaum, Paul R. 2010. *Design of Observational Studies.* New York: Springer.

Rosenbaum, Paul R., and Donald B. Rubin.1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55.

Rosenthal, Robert, and Lenore Jacobson. 1968. *Pygmalion in the classroom.* New York: Holt, Rinehart & Winston.

Sampson, Robert J., and John H. Laub. 1997. A Life-Course Theory of Cumulative Disadvantage and the Stability of Delinquency. In *Developmental Theories of Crime and Delinquency*, Advances in Criminological Theory, Volume 7, ed. T. P. Thornberry. New Brunswick, NJ: Transaction.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. Neighborhoods and violent crime: A multilevel study of collective efficacy. Science 277:918–24.

Schur, Edwin W. 1973. *Radical Non-intervention: Rethinking the Delinquency Problem*. Englewood Cliffs, NJ: Prentice-Hall.

Schwartz, Richard, and Jerome Skolnick. 1962. Two studies of legal stigma. Social Problems 10:133–42.

Short, James F. Jr. 1958. Extent of unrecorded juvenile delinquency: Tentative conclusions.

Journal of Criminal Law and Criminology 49:296–302.

Short, James F. Jr., and F. Ivan Nye. 1957. Reported behavior as a criterion of deviant behavior. Social Problems 5:207–13.

Simon, Jonathan. 2007. *Governing through Crime: How the War on Crime Transformed American Democracy and Created a Culture of Fear*. New York: Oxford University Press.

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.

Snyder, Howard N., and Melissa Sickmund. 2006. *Juvenile Offenders and Victims: 2006 National Report*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Stuart, Elizabeth A. 2010. Matching methods for causal inference: A review and a look forward. Statistical Science 25:1–21.

Stucki, Bryce. 2013. Reversing Broward County's School-to-Prison Pipeline. *American Prospect*, 4 December. Available: http://prospect.org/article/reversing-broward-countys-school-prison-pipeline

Sweeten, Gary. 2006. Who will graduate? Disruption of high school education by arrest and court involvement. Justice Quarterly 23:462–80.

Tannenbaum, Frank. 1938. *Crime and the Community*. New York and London: Columbia University Press.

Uggen, Christopher. 2000. Work as a turning point in the life course of criminals: A duration model of age, employment, and recidivism. American Sociological Review 65:529–46.

U.S. Department of Education. 2014. *Guiding Principles: A Resource Guide for Improving*

School Climate and Discipline. Washington, DC: U.S. Department of Education. Available: http://www2.ed.gov/policy/gen/guid/school-discipline/guiding-principles.pdf

U.S. Office of Juvenile Justice and Delinquency Prevention. 2012. *OJJDP Statistical Briefing Book*. Online. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention. Available: http://www.ojjdp.gov/ojstatbb/crime/JAR_Display.asp?ID=qa05201.

van Buuren, S., J. P. L. Brand, C. G. M. Groothuis-Oudshoorn, and D. B. Rubin. 2006. Fully conditional specification in multivariate imputation. Journal of Statistical Computation and Simulation 76: 1049–1064.

Wallerstein, J. S., and C. J. Wylie. 1947. Our law-abiding law-breakers. Probation 25:107–12.

Wiley, Stephanie A., and Finn-Aage Esbensen. In press. The effect of police contact: Does official intervention result in deviance amplification? Crime and Delinquency.

Wiley, Stephanie A., Lee Ann Slocum, and Finn-Aage Esbensen. 2013. The unintended consequences of being stopped and arrested: An exploration of the labeling mechanisms through which police contact leads to subsequent delinquency. Criminology 51:927–66.

Table 1. Differences between Arrestees and Non-arrestees, Before and after Matching: Individual-level Factors.

	UNMATCHED					MATCHED				
Wave 1 Variable: Individual Factors	Arrested Mean	Not Arrested Mean	t-test	sig	% absolute bias	Arrested Mean	Not Arrested Mean	t-test	sig	% absolute bias
SRO & Aggression										
Violent offending	0.60	0.15	9.30	*	84.9	0.45	0.42	0.17		4.5
Property offending	0.22	0.09	4.54	*	45.3	0.15	0.16	-0.15		3.4
Drug selling	0.14	0.02	5.78	*	41.4	0.07	0.06	0.10		2.0
Public order offending	0.21	0.11	2.68	*	32.2	0.18	0.26	-0.94		26.0
Status offending	0.64	0.12	10.13	*	97.9	0.42	0.47	-0.38		9.1
Variety score	4.28	1.80	7.42	*	82.5	3.50	3.40	0.13		3.2
Delinquent score (Achenbach)	4.98	3.24	4.34	*	56.1	4.34	3.97	0.53		12.1
Aggressive score (Achenbach)	9.83	8.83	1.16		15.0	9.61	9.74	-0.08		2.0
Temperament										
Inhibitory control	2.80	2.41	2.94	*	39.1	2.68	2.68	0.01		0.1
Impulsivity	3.19	2.98	1.87		25.8	3.06	3.01	0.25		6.0
Sensation seeking	3.07	2.71	3.33	*	47.7	2.93	2.84	0.52		12.5
Activeness	3.80	3.57	1.77		26.2	3.70	3.47	1.12		26.5
Emotionality	2.81	2.67	0.93		12.9	2.74	2.58	0.64		14.5
Sociability	3.69	3.64	0.48		7.3	3.66	3.61	0.28		6.6
Shyness	2.25	2.48	-1.91		28.5	2.17	2.47	-1.64		35.6
School & Education										
Ever repeated grade	0.22	0.15	1.30		18.3	0.25	0.18	0.64		16.8
Ever remediation class	0.33	0.29	0.58		8.6	0.31	0.32	-0.03		0.8
School truancy	0.50	0.16	4.37	*	50.1	0.30	0.45	-0.81		22.1
IQ (standardized)	96.89	100.78	-1.83		28.8	97.86	98.94	-0.34		8.0
school mobility	3.13	2.48	3.69	*	49.9	2.82	2.74	0.27		5.7
Drug Use										
days marij. last month	1.30	1.11	2.05	*	23.4	1.21	1.12	0.74		10.8

days cigarettes last month	1.64	1.22	3.07	*	33.5	1.58	1.43	0.47	11.7
days alcohol last month	1.25	1.12	1.88		25.3	1.24	1.12	0.93	23.2
Demographics									
Gender	0.72	0.45	3.75	*	55.1	0.66	0.63	0.28	6.4
Age	0.51	-0.55	5.00	*	75.1	0.27	0.07	0.58	13.9
African American	0.64	0.34	4.57	*	63.7	0.63	0.66	-0.24	5.5
Mexican	0.17	0.32	-2.32	*	35.7	0.18	0.18	0.00	0.0
Puerto Rican or other Hispanic	0.08	0.14	-1.28		19.9	0.08	0.10	-0.27	5.7
Other race	0.06	0.03	0.89		11.0	0.03	0.02	0.26	4.2
1st generation immigrant	0.09	0.15	-1.07		16.2	0.11	0.07	0.53	10.8
2nd generation immigrant	0.15	0.31	-2.42	*	37.7	0.13	0.10	0.48	8.5
N	53	951				38	111		

** significant at p<.05.*

Table 2. Differences between Arrestees and Non-arrestees, Before and after Matching: Family Factors.

	UNMATCHED					MATCHED			
Wave 1 Variable: Family Factors	Arrested Mean	Not Arrested Mean	t-test			Arrested Mean	Not Arrested Mean	t-test	% absolute bias
Household Composition									
Parents married	0.23	0.54	-4.45	*	67.3	0.24	0.24	0.00	0.0
Extended family in household	0.26	0.19	1.28		17.1	0.21	0.31	-0.95	22.9
# children	3.55	3.36	0.77		10.0	3.29	3.24	0.10	2.5
Single parent	0.45	0.31	2.17	*	29.5	0.47	0.54	-0.53	12.7
SES & Residence									
Years current address (subject)	5.78	5.87	-0.14		1.8	5.22	6.88	-1.37	33.5
Years current address (primary caregiver)	7.78	6.98	0.76		9.8	6.64	8.67	-1.09	24.9
Caregiver occupational status (SEI)	41.42	42.29	-0.35		5.2	41.26	39.23	0.56	12.1
Caregiver education	3.08	3.04	0.21		2.9	3.03	2.92	0.37	8.1
Household income	3.79	4.25	-1.73		25.7	3.82	3.85	-0.09	2.0
SES composite	-0.22	-0.06	-0.81		11.9	-0.24	-0.34	0.34	7.2
Home Interior physical envir.	-0.25	0.05	-1.14		16.0	-0.24	-0.36	0.30	6.8
Home exterior physical envir.	-0.28	0.04	-1.72		25.4	-0.21	-0.52	1.04	24.3
Parent Risk									
Father criminal involv.	0.09	0.11	-0.45		6.6	0.11	0.14	-0.46	11.4
Father substance use	0.21	0.14	1.29		16.9	0.24	0.26	-0.26	6.9
Mother substance use	0.08	0.05	0.88		11.2	0.05	0.11	-0.96	25.4
Mother depression	0.19	0.14	0.94		12.5	0.16	0.09	0.92	18.8
Family and Parenting Processes									
Family supervision	-0.45	-0.05	-3.67	*	47.6	-0.29	-0.13	-0.86	18.3
Parent-child conflict	0.27	-0.07	3.21	*	41.2	0.19	0.25	-0.29	7.2
Family conflict	51.08	47.48	2.49	*	32.5	49.24	50.00	-0.32	6.9
Disciplined child-rearing	58.36	58.31	0.04		0.5	58.26	58.78	-0.29	6.1

Religiosity score	62.77	60.44	2.18	*	35.8	62.94	62.22	0.53	11.2
Family support	-0.12	0.01	-1.14		15.7	0.08	-0.06	0.70	16.0
Lack of hostility (primary caregiver)	-0.15	0.34	-0.79		11.9	-0.07	-0.26	0.22	4.5
Developmental Environment									
Access to reading	-0.38	-0.03	-1.32		17.9	-0.22	-0.07	-0.35	8.0
Developmental stimulation	-0.48	-0.07	-2.87	*	36.2	-0.38	-0.03	-1.46	31.5
Family outings	-0.10	-0.03	-0.59		7.9	-0.03	-0.09	0.35	7.8
Parental verbal ability	-0.26	0.06	-1.28		16.6	-0.01	0.03	-0.11	2.2
Parental warmth	-0.55	0.04	-2.53	*	33.8	-0.48	-0.50	0.04	0.8
N	38	951				38	111		

** significant at $p < .05$.*

Table 3. Differences between Arrestees and Non-arrestees, Before and after Matching: Peer and Neighborhood Factors.

	UNMATCHED					MATCHED			
Wave 1 Variable	Arrested Mean	Not Arrested Mean	t-test			Arrested Mean	Not Arrested Mean	t-test	% absolute bias
Peer Factors									
Peer pressure	0.47	-0.06	3.55	*	46.6	0.47	0.50	-0.10	2.5
Peer attachment to school	0.11	-0.01	1.98	*	25.7	0.04	0.00	0.37	7.6
Friend support	0.06	0.06	0.04		0.6	0.07	0.02	0.39	9.3
Deviance of peers	0.55	-0.05	5.33	*	72.0	0.42	0.35	0.37	8.3
Peer attachment	-0.18	0.07	-2.55	*	34.3	-0.09	-0.21	0.68	16.9
Neighborhood Factors									
Legal cynicism	2.55	2.50	2.73	*	33.3	2.53	2.55	-0.71	17.3
Tolerance of deviance	4.24	4.24	0.14		2.1	4.24	4.25	-0.25	5.6
Perceived social disorder	2.13	2.00	2.54	*	36.2	2.09	2.13	-0.46	10.4
Perceived physical disorder	1.74	1.65	2.34	*	31.4	1.71	1.75	-0.56	12.8
Neighborhood organizations	-0.29	-0.41	1.53		22.1	-0.30	-0.25	-0.44	9.9
Services for youth	-1.63	-1.74	1.08		14.7	-1.68	-1.55	-0.78	18.0
Collective efficacy	3.84	3.89	-1.49		20.4	3.87	3.84	0.47	10.6
Residential stability	0.06	-0.01	0.46		6.3	0.13	0.03	0.43	10.2
Resident victimization last 6 mos.	0.39	0.42	-1.25		18.1	0.36	0.37	-0.17	3.6
% foreign born	13.44	21.28	-3.66	*	52.4	13.03	12.37	0.20	4.4
Concentrated poverty	0.28	-0.14	4.12	*	52.0	0.22	0.44	-0.97	26.7
Concentrated affluence	-0.28	-0.21	-0.81		11.7	-0.26	-0.32	0.35	7.8
1995 official violent crime	9.20	8.85	3.98	*	57.3	9.11	9.27	-1.05	25.2
% black	55.61	31.37	4.60	*	62.4	53.95	57.79	-0.40	9.9
% Latino	22.13	32.11	-2.41	*	34.3	20.91	20.91	0.00	0.0
N	53	951				38	111		

** significant at p< .05.*

Table 4. Rosenbaum Bounds, Effect of First Arrest

	Future Arrest		Violent Offending	
Γ	Q+	p-value	Q+	p-value
1.00	4.320	<.001	2.550	0.005
1.25	3.720	<.001	1.992	0.023
1.50	3.236	0.001	1.533	0.063
1.75	2.836	0.002	1.148	0.125
2.00	2.496	0.006	0.817	0.207
2.25	2.200	0.014	0.526	0.299
2.50	1.938	0.026	0.266	0.395
2.75	1.703	0.044	0.031	0.488
3.00	1.490	0.068	-0.183	0.573

Note: Γ refers to the odds ratio of the effect of unobserved variables on the likelihood of first arrest for youths who were arrested versus youths who were not arrested.

Table 5. Effects of First Arrest on Rearrest

	Model 1					Model 2					Model 3				
	Coef	SE	OR	t	sig	Coef	SE	OR	t	sig	Coef	SE	OR	t	sig
First Arrest	1.60	0.41	4.97	3.92	***	1.63	0.41	5.10	3.98	***	1.40	0.55	4.06	2.55	*
Wave 3 Violent Offending						-0.21	0.44	0.81	-0.48		-0.58	0.54	0.56	-1.07	
Arrest x Offending											0.55	0.81	1.73	0.68	
Constant	-1.33	0.27	0.26	-4.87	***	-1.24	0.32	0.29	-3.86	***	-1.11	0.33	0.33	-3.41	***
N	581					581					581				

ABBREVIATIONS: Coeff = b coefficient; SE = standard error; OR = odds ratio; sig = significance.

Significance: *** p < .001; ** p <.01; * p < .05; ~ p<.10 (two-tailed tests).

Table A-1. Alternative Specifications

	Means		Tests					Absolute Bias	
Specification	Arrestees	Controls	Diff or OR	SE	t or z	Sig	Treatment *N* On Support	Mean	Median
Arrest (logit)									
NN, 3:1, caliper 0.02	0.49	0.16	5.20	1.12	4.63	***	43/58	6.9	4.8
NN, 1:1, caliper 0.02	0.48	0.14	5.45	1.74	3.14	**	42/58	11.8	9.5
NN, 3:1, caliper 0.04	0.48	0.14	5.41	1.11	4.86	***	46/58	7.5	6.8
KM, BW = .06	0.51	0.13	7.18	1.33	5.41	***	49/58	8.7	6.8
KM, BW = .10	0.51	0.12	7.45	1.31	5.68	***	49/58	12.1	10.7
Variety (continuous)									
NN, 3:1, caliper 0.02	2.03	1.04	0.98	0.40	2.44	*	38/53	11.1	8.5
NN, 1:1, caliper 0.02	1.89	0.94	0.94	0.44	2.13	*	36/53	13.7	11.6
NN, 3:1, caliper 0.04	1.88	1.18	0.70	0.39	1.79	+	41/53	11.5	10.0
KM, BW = .06	1.84	1.62	0.22	0.37	0.58		44/53	12.0	10.5
KM, BW = .10	1.84	1.41	0.43	0.35	1.21		44/53	14.1	12.6
Violence (logit)									
NN, 3:1, caliper 0.02	0.58	0.30	3.24	1.24	2.61	**	38/53	11.1	8.5
NN, 1:1, caliper 0.02	0.56	0.28	3.25	1.53	2.12	*	36/53	13.7	11.6
NN, 3:1, caliper 0.04	0.54	0.34	2.23	1.28	1.75	+	41/53	11.5	10.0
KM, BW = .06	0.52	0.37	1.86	1.21	1.53		44/53	12.0	10.5
KM, BW = .10	0.52	0.33	2.26	1.04	2.17	*	44/53	14.1	12.6
Property (logit)									
NN, 3:1, caliper 0.02	0.29	0.16	2.17	1.26	1.72	+	38/53	11.1	8.5
NN, 1:1, caliper 0.02	0.28	0.11	3.08	1.65	1.87	+	36/53	13.7	11.6
NN, 3:1, caliper 0.04	0.27	0.15	2.14	1.24	1.72	+	41/53	11.5	10.0
KM, BW = .06	0.27	0.19	1.63	1.37	1.19		44/53	12.0	10.5
KM, BW = .10	0.27	0.18	1.67	1.24	1.34		44/53	14.1	12.6
Drug Selling (logit)									
NN, 3:1, caliper 0.02	0.13	0.09	1.49	2.53	0.59		38/53	11.1	8.5
NN, 1:1, caliper 0.02	0.11	0.08	1.38	3.72	0.37		36/53	13.7	11.6
NN, 3:1, caliper 0.04	0.12	0.11	1.18	4.90	0.24		41/53	11.5	10.0
KM, BW = .06	0.11	0.16	0.66	-1.10	-0.60		44/53	12.0	10.5
KM, BW = .10	0.11	0.12	0.91	-6.06	-0.15		44/53	14.1	12.6

NOTES: Matching for SRO outcomes utilizes 53 arrestees and 951 nonarrestees; matching with arrest outcomes utilizes 58 arrestees and 1191 nonarrestees. Binary outcomes estimated via logit models with a Z-test of significance; continuous outcomes estimated via psmatch2 with a t-test of significance.

ABBREVIATIONS: NN = Nearest Neighbor; KM = Kernel Matching; BW = bandwidth; sig = significance.

+p < .10; * p < .05; ** p < .01; *** p < .001 (two-tailed tests).

Figure 1. The Distribution of Propensity Scores of First Arrest, By Treatment Status.

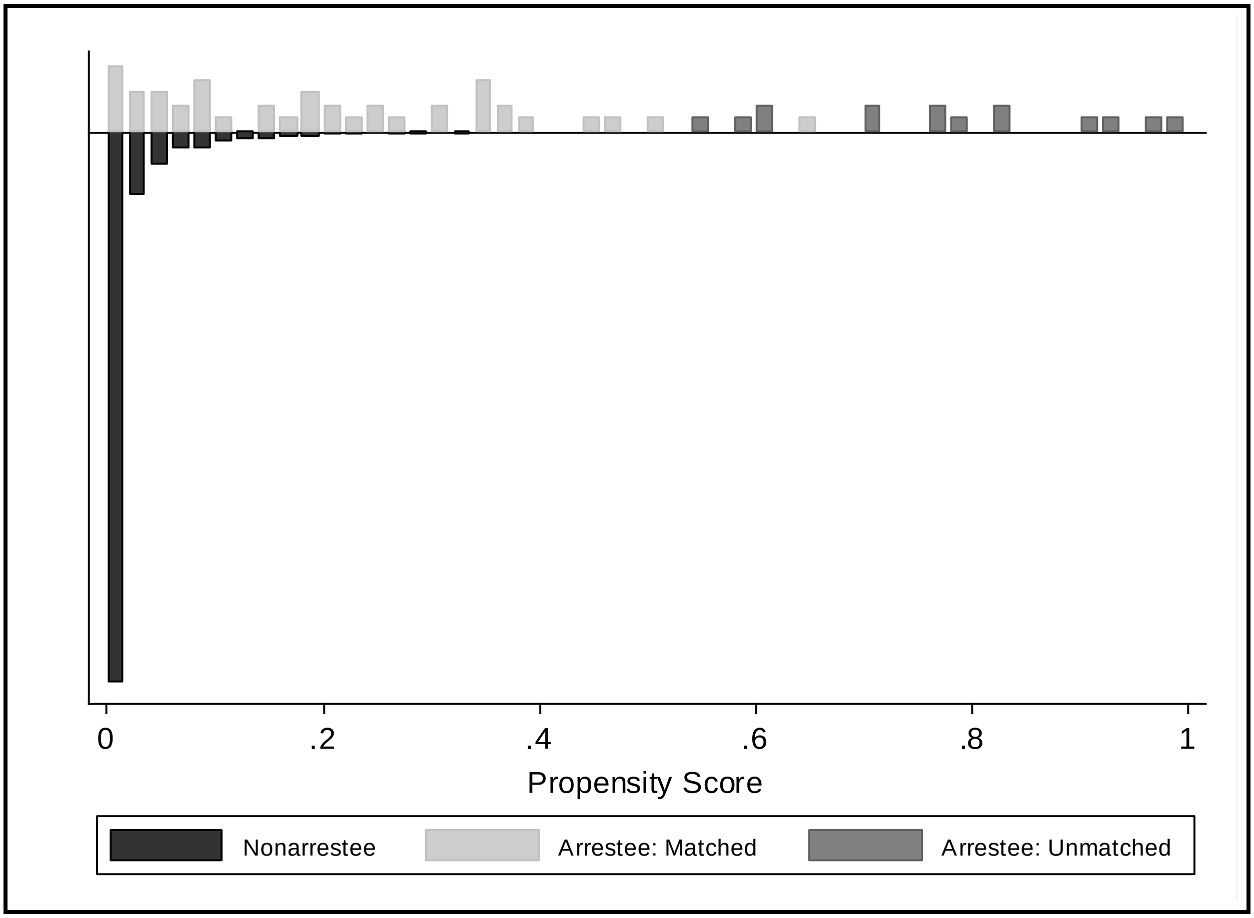
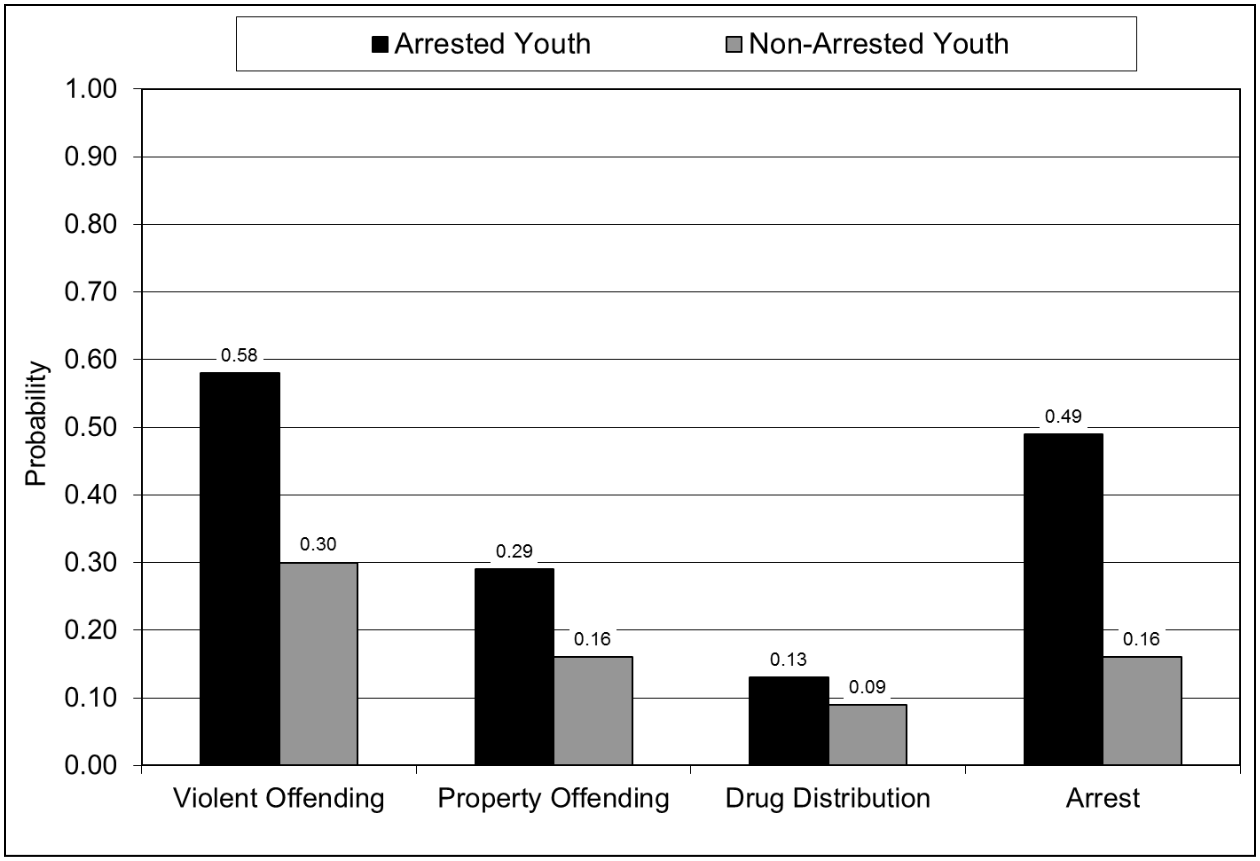


Figure 2: The Effect of First Arrest of the Probability of Subsequent Offending and Rearrest, Individually Matched Arrested and Non-Arrested Youths



APPENDIX A: METHODOLOGY

BOUNDS FOR THE EFFECT OF ARREST ON RE-ARREST

We use Rosenbaum's (2002) bounding approach to examine the sensitivity of our propensity-matched inferences to hidden biases (see also Becker and Caliendo, 2007; DiPrete and Gangl, 2004). This approach allows us to determine how strongly an omitted confounding variable must influence selection into treatment to undermine our inferences about the causal effect of arrest. If there is hidden bias, then two individuals with the same observed characteristics will have differing likelihoods of being arrested because of unobserved factors.

The odds that an individual will receive treatment is given by the following:

$$\frac{\Pr(Arrest = 1)}{1 - \Pr(Arrest = 1)} = \exp(\alpha + \beta X + \gamma U),$$

where X represents observed variables and U represents unobserved variables. In this case, the variable U increases the probability of arrest by a factor equal to γ. For a pair of individuals i and j matched on propensity score (i.e., the same observed covariates X), where i is ultimately arrested and j is not, the ratio of odds of receiving treatment is given by:

$$\frac{\frac{P_i}{1 - P_i}}{\frac{P_j}{1 - P_j}} = \frac{\exp(\alpha + \beta X_i + \gamma U_i)}{\exp(\alpha + \beta X_j + \gamma U_j)}$$

Because i and j have the same set of observed covariates, X cancels out:

$$\frac{\exp(\gamma U_i)}{\exp(\gamma U_j)} = \exp\left[\gamma\left(U_i - U_j\right)\right]$$

If there are no differences in unobserved variables (Ui = Uj for all matched pairs) or if unobserved variables have no influence on the probability of arrest (γ=0), then there is no hidden bias. Since we do not have direct information on unobservables, we use a sensitivity analysis to evaluate whether our statistical inferences pertaining to the effect of arrest on subsequent arrest would change under different values of γ. Per Rosenbaum (2002), the bounds on the odds ratio that either of the two matched individuals will receive treatment is given by:

$$\frac{1}{e^{\gamma}} \leq \frac{P_i(1-P_j)}{P_j(1-P_i)} \leq e^{\gamma},$$

where $\Gamma=\exp(\gamma)$. Use of this bounding approach is suitable if pairwise matching is done without replacement (Becker and Caliendo, 2007).

We use the mhbounds routine in Stata to implement our sensitivity analysis. The mhbounds command uses the Mantel and Haenszel (MH; 1959) test statistic. The Q+ test-statistic adjusts the MH statistic downward in the event of positive unobserved selection. Positive selection occurs when arrested individuals are more likely to be arrested again in the future for reasons other than their prior arrest. In this case, we would overestimate the treatment effect of arrest.

PROPENSITY SCORE WEIGHTS

Rather than using propensity scores for matching, for our analysis of the effect of arrest on future arrest net of offending, we use propensity scores as inverse weights in order to estimate the average treatment effect on the treated (Hirano and Imbens, 2001; Hirano, Imbens, and Ridder, 2003; Stuart, 2010). In equation form, the weight equals:

$$w_i = T_i + (1-T_i)\frac{\hat{e}_i}{1-\hat{e}_i},$$

where $\hat{e}_i$ is the estimated propensity score for individual i. Per this formula, treated individuals (i.e., arrestees) receive a propensity weight equal to one. Control individuals with greater propensity scores receive a larger weight. One particular advantage of using propensity score weights is that we can use the weights in a regression model that includes covariate adjustment. In our case, per Hypothesis 3, we are interested in estimating the effect of arrest on subsequent arrest after adjusting for the extent of criminal offending at Wave 3 of the PHDCN.

ALTERNATIVE PROPENSITY SCORE SPECIFICATIONS

Along with the 3:1 nearest neighbor (NN) matching approach already described, we estimated effects through four other propensity-score-matching specifications. We used two alternative specifications with NN matching: an expanded caliper (0.04) in 3:1 matching, and 1:1 matching with caliper=0.02. We also explored two kernel matching (KM) specifications. In contrast to NN, which uses just a few observations from a potential control group to create a counterfactual, KM uses a weighted average of all non-arrestees to construct the counterfactual outcome (Caliendo and Kopeinig, 2008; Stuart, 2010). One advantage of kernel matching is reduced variance, because more observations and therefore more information are used to construct the counterfactual. The tradeoff for reduced variance is an increase in bias. KM has similarities with the inverse probability of treatment weighting previously described, with the main distinction being the type of weighting employed (i.e., kernel weighting versus an inverse of the probability).

Results are shown in Table A-1. The approach already described (3:1 nearest neighbor matching, caliper = 0.02) was the most efficient in removing bias. At the same time it was more stringent than most of the other specifications on matching, so that 38 of 53 arrestees with SRO data were matched (a.k.a., "on support"), and 43 of 58 with future arrest data. The other specifications were somewhat less efficient in reducing bias, but were more liberal in retaining matching controls to arrestees.