THE EFFECTS OF DNA DATABASES ON CRIME

Jennifer L. Doleac*

August 2016

Abstract

Every U.S. state has a database of criminal offenders' DNA profiles. These databases receive widespread attention in the media and popular culture, but there has been no rigorous analysis of their impact on crime. This paper intends to fill that gap. I exploit the details and timing of state DNA database expansions in two ways, first to address the effects of DNA profiling on individuals' subsequent criminal behavior and then to address the aggregate effects on crime rates. I show that DNA databases deter crime by profiled offenders, reduce crime rates, and are more cost-effective than traditional law enforcement tools.

JEL Classifications: K14, K42, H7

I am grateful to Ran Abramitzky, Doug Bernheim, David Bjerk, David Eil, Bill Evans, Leora Friedberg, Bill Gale, Brandon Garrett, Caroline Hoxby, Ilyana Kuziemko, Jonathan Meer, Nicholas Sanders, Kaitlin Shilling, and several anonymous reviewers for helpful suggestions and comments. Thanks also to seminar participants at the following: American Enterprise Institute, Brookings Institution, College of William & Mary, Harvard Kennedy School, University of Maryland, University of Notre Dame, University of Pennsylvania Law School, RAND, Stanford University, Tulane University, University of Virginia, Wellesley College, NBER Summer Institute Crime Working Group, and IRP Summer Research Workshop. I appreciate the financial support of the Hawley-Shoven Fellowship and the John M. Olin Program in Law and Economics, both at Stanford University.

*Frank Batten School of Leadership and Public Policy, University of Virginia, Charlottesville, VA 22904. Email: jdoleac@virginia.edu.

1 Introduction

Technology is transforming our criminal justice system. Of particular interest are new tools that make it easier and less expensive to identify criminal offenders, often through increased surveillance. These provide an important alternative to traditional methods of crime-prevention, particularly long-term incarceration. However, because these tools are so different from previous crime-prevention methods, their effects are uncertain.

In this paper, I analyze the effects of one high-tech law enforcement tool: DNA databases. Beginning in 1988, states passed legislation to create (and subsequently expand) databases of criminal offenders' DNA profiles. The goal of cataloguing these genetic fingerprints is to quickly and accurately match known offenders with crime scene evidence, thereby deterring profiled offenders from committing more crime and taking serial offenders off the streets more quickly. This paper addresses two questions related to databases' effectiveness: How has DNA profiling affected profiled offenders' criminal behavior, and what are the databases' aggregate effects on crime rates?

State-level legislation periodically expands state databases to add new categories of offenders; these expansions provide useful variation that allows me to address both questions using quasi-experimental methods. By comparing offenders released before and after the laws' effective dates, I can measure the individual-level effect on profiled offenders' behavior. Individuals released just after those dates are substantially more likely to be required to provide a DNA sample, but otherwise look very similar to those released just before. Subsequent differences in behavior can thus be attributed to the effect of the DNA requirement. Using criminal history data from seven states, I compare recidivism rates for offenders released pre- and post-expansion to test whether DNA profiling has a deterrent effect on crime.

Next, I use annual state-level data on DNA database size and crime rates from 2000 to 2010 to test the aggregate effect of DNA profiling on public safety. Database size – a proxy for the probability that a would-be offender is in the database – is endogenous, so I use the timing of the above-mentioned expansions to construct instrumental variables. The timing of the expansions is unrelated to underlying crime trends or state characteristics that might affect future crime rates, and so these law changes provide exogenous shocks to the number of profiles in a state database. Using these instruments, I measure the effect of database size on state crime rates.

In these analyses, I focus on database expansions that targeted felony convicts and inmates – the relevant groups during this period – and find significant crime-reducing effects on both the individual and aggregate levels.

At the individual level, I first show that the pre- and post-expansion groups look very similar in terms of their propensity to reoffend. I then estimate that the requirement to submit a DNA sample reduces the likelihood of a new conviction within five years by 4.5

percentage points (17%) for serious violent offenders.[1] This effect is statistically significant ($p < 0.05$) and highly robust to alternative samples and specifications. I also estimate that the DNA requirement reduces 5-year recidivism by serious property offenders, by 2.4 percentage points (6%). This effect is only marginally significant ($p < 0.10$) and less robust, but provides suggestive evidence that DNA profiling also affects the behavior of this group. Given that DNA profiling should also have a positive probative effect – conditional on offending, it increases the probability of getting caught – the magnitudes of these estimates represent lower bounds on the true deterrent effects. These results are consistent for different choices of sample selection criteria and functional forms, which I present in Section 3.4.

While these effects on profiled offenders are encouraging, the aggregate effect on crime is of greater relevance for policy-makers. Both deterrence and incapacitation should decrease the total amount of crime in each state, barring rapid replacement by new offenders. As more potential reoffenders are added to state DNA databases, the number of crimes should fall. Does it?

OLS estimates of the effect of database size on crime rates will be biased upwards because the number of profiles and number of crimes in a state are simultaneously determined. Also, state governments' adeptness and degree of motivation with regard to implementing database expansions affect the number of profiles uploaded. These state characteristics might also affect crime rates, resulting in omitted variable bias. I use an instrumental variable (IV) approach to eliminate these biases.

Specifically, I take advantage of substantial, exogenous variation in the timing of state database expansions. I first show that these expansion dates are not correlated with pre-period crime rates or state characteristics that might independently affect crime trends. I then exploit this variation to construct a simulated IV that predicts the number of qualifying offenders over time in each state, based on the timing of expansions and pre-period conviction rates. This instrument estimates the number of profiles that *should* be in each database, rather than the number that are actually uploaded. The resulting instrument is highly correlated with actual database size, but is not simultaneously determined or affected by how well states implement their database laws. It is correlated with crime rates *only* through its correlation with database size.

Using this simulated instrument and data I collected on database size in each state, I find that, between 2000 and 2010, increasing the size of state databases lowered crime rates. Based on my preferred specification, one additional DNA profile per 100,000 residents decreased violent crime by 0.051 offenses per 100,000 residents per year, and decreased property crime by 0.323 offenses per 100,000 residents per year. These effect sizes are large, but imprecisely measured. The 95% confidence intervals imply that growth in the average database from 2000 to 2010 decreased violent crime by 7–45% and property crime by 5–

[1] I will use the term "serious offense" to refer to FBI Index I offenses: felony homicide and non-negligent manslaughter, forcible rape, aggravated assault, robbery, burglary, larceny, vehicle theft. and arson.

35%. I find statistically significant, negative effects on each individual type of crime, except for burglary. The absence of any significant impact on burglary rates may be due to more limited use of DNA evidence from property crime scenes[2], and limited crossover between burglary and other crimes. It might also imply a high replacement rate for burglary.[3]

These results are robust to alternative IV specifications, and there is no evidence that pre-existing trends explain the effects. Database size does not have a statistically-significant effect on arrest rates, providing more evidence that databases' effects on crime are driven by deterrence, not incapacitation. Finally, I do not find clear evidence that the benefit of adding additional profiles increases or decreases as databases expand. In other words, adding minor felons appears to have the same crime-reducing effect as adding serious felons.

DNA profiling reduces crime, but is it worth the cost? Unlike law enforcement tools such as prisons and police officers, DNA databases exhibit tremendous returns to scale. Initial investments in computer infrastructure and crime labs were large, but the marginal cost of adding an additional offender to an existing database is low (currently less than \$40), and this cost is falling rapidly as technology improves. Given the low marginal cost of a DNA profile, databases must only decrease crime a small amount to justify the financial cost of database expansions. (Privacy concerns are a separate issue and clearly more difficult to quantify.) Based on others' estimates of the effects of sentence enhancements and police hiring on crime rates, I calculate that the marginal cost of preventing a serious offense is about \$7,600 using longer sentences and \$26,300-62,500 using police officers. In contrast, my results on the effect of DNA databases suggest that the marginal cost of preventing a serious offense using DNA profiling is under \$600, and falling.

This paper contributes to the literature in the following ways: (1) It is the first attempt to measure the effect of DNA databases on crime, and the first to convincingly estimate the effect of DNA profiling on criminal behavior. Given the widespread use of this technology around the world, and ongoing public debates regarding the use and expansion of existing databases, the findings of this study are highly policy-relevant. (2) These findings speak to the likely effects of other high-tech law enforcement tools, which are frequently adopted by law enforcement but rarely evaluated. (3) This study fits into the economics of crime literature on how changing the probability of punishment, rather than punishment severity, affects criminal behavior. In particular, it provides evidence that even serious offenders respond to a higher probability of getting caught by committing fewer offenses. (4) This paper speaks to to the larger issue of how we interact with technology, and how government surveillance affects our behavior.

I proceed as follows: Section 2 provides background on DNA database policy in the

[2]Though useful DNA evidence is usually available at burglary crime scenes, it is typically not collected or analyzed because violent crimes are given higher priority by crime labs.

[3]That is, as profiled offenders are deterred or incapacitated, new, unprofiled offenders quickly take their place.

United States; section 3 discusses the data, empirical strategy, and results for the individual-level analysis; section 4 does the same for the crime rate analysis; section 5 considers the cost effectiveness of DNA databases; and section 6 concludes.

2 Background on DNA database policy in the U.S.

In 1988, Colorado began collecting some convicted sex offenders' DNA; Virginia followed suit in 1989, requiring blood samples from sex and violent offenders, then expanding its law to include all convicted felons in 1990. Soon, DNA profiling required a simple saliva swab instead of a blood sample, decreasing the cost and invasiveness of DNA collection. Taking either a blood or saliva sample is quite salient to offenders, and it appears that most were well aware of the purpose of DNA collection.[4]

By 1999, with the urging and financial support of the federal government, every state had established an offender DNA database. The FBI currently links all states' databases to form the Combined DNA Index System (CODIS), which contained 12.1 million offender profiles as of December 2015[5]; this national database is well-known to the American public thanks to popular television crime dramas like *CSI*. However, there are financial and privacy costs from collecting and analyzing offenders' DNA samples, and – while anecdotal successes are widely touted – it has been unclear whether the databases have public safety benefits.

The goal of these databases was not to be tougher on criminals, per se, but to increase accuracy and hold the right people accountable for their crimes.[6] They appealed to legislators as much for their potential to exonerate wrongly-convicted offenders (and prevent new mistakes) as they did for their ability to lock up career criminals.[7] For this reason, liberal

[4]There is anecdotal evidence that law enforcement emphasized the purpose of DNA databases at the time of collection. Many offenders tried to avoid providing samples, suggesting they knew the probative power of DNA profiling. (Resulting legal challenges were defeated in every state.)

[5]In addition, CODIS contains 2.2 million DNA profiles from arrestees. While the addition of arrestees to state databases is an important policy issue, this paper focuses on the addition of convicted offender profiles only. Current statistics are available at `http://www.fbi.gov/about-us/lab/biometric-analysis/codis/ndis-statistics`.

[6]Racial bias within the criminal justice system is a primary concern (Ridgeway and MacDonald, 2009). Tools like DNA databases have the potential to decrease racial bias by objectively identifying repeat offenders—and rejecting incorrect identifications by police and eyewitnesses—without regard to race.

[7]Showing that DNA evidence does not match a convicted offender is often not enough to exonerate him in practice. In an interview with the Council for Responsible Genetics, Peter Neufeld, co-founder of the Innocence Project, described how DNA databases help exonerate wrongly-convicted individuals: "There are occasions where we get a DNA test result on a material piece of evidence from a crime scene which would exclude our client, but prosecutors still resist motions to vacate the conviction. In some of those cases, what then tipped the balance in our favor was that the profile of the unknown individual [whose DNA was found at the crime scene] was run through a convicted offender database and a hit was secured. Once we were able to identify the source of the semen or blood... we were then able to secure the vacation of the conviction for our client." He went on to add, "There's no question that there would be fewer wrongful convictions if there was a universal DNA databank." (CRG Staff, 2011) In 2007, Barry Scheck, the other co-founder of the Innocence Project, told the New York Times that "many of the people his organization had helped exonerate would have been freed much sooner, or would not have been convicted at all" if state databases included profiles from all convicted offenders. (McGeehan, 2007)

states were as likely as conservative states to quickly add new qualifying offenses.

However, once the laws went into effect, states varied in their ability to promptly collect and analyze DNA samples from qualifying offenders. Collection was relatively quick and inexpensive, and therefore began on time, but the rate at which these samples were converted into searchable profiles (that is, analyzed in a laboratory and uploaded to the database) varied by state. The result was often large backlogs of samples waiting for analysis.[8]

This imperfect implementation is important for two reasons: First, I observe when a DNA sample was required, not when a sample was taken and uploaded. It is possible that authorities did not collect DNA samples from some of the offenders who were required to provide one, or that analysis was delayed. The individual-level analysis therefore measures the effect of the intention to treat (DNA requirement), not of the actual treatment (DNA profiling). Second, database size is likely a function of states' adeptness and motivation in analyzing DNA profiles, and these characteristics might affect crime through other channels. I use an IV strategy in the crime rate analysis to correct for any omitted variable bias.

This paper focuses on the effect of profiling convicted felons, and most states included all felony convicts in their databases by 2010. However, the use and further expansion of DNA databases are ongoing policy issues in the United States and abroad. States continue to add new groups of individuals, particularly misdemeanor convicts and felony arrestees.[9]

2.1 Related literature

Becker (1968) hypothesized that fewer people will choose to commit crime when the expected punishment increases. That is, an individual will only offend if:

$$\text{E(Benefit)} > \text{E(Cost)} \text{ and } \text{I(Incarcerated)} = 0, \tag{1}$$

where E(Cost) = f(p, δ, s); p is the probability of conviction, conditional on reoffending; δ is a discount factor; and s is the punishment (e.g., sentence length). E(Cost) is increasing in each of these parameters. I(Incarcerated) indicates whether the individual is currently incarcerated.

Consider a recently-released criminal offender who is deciding whether to commit another crime. We can decrease the probability that he reoffends by increasing p or s; this is the deterrent effect. If he reoffends and is convicted of the crime, which happens with probability p, he will be unable to reoffend again because he is in prison; this is the incapacitation effect.

[8] It is extremely unlikely that offenders were aware of the size of sample backlogs; even policymakers were unaware of this problem for years. If they were aware of the delays in analyzing DNA samples, this would decrease the deterrent effect and bias my results toward zero.

[9] The Supreme Court's recent decision in *Maryland v. King* — establishing that DNA collection from arrestees is constitutional — drew attention to this issue and incited many voters and legislators to reconsider their states' laws. For more information on this case, see `http://www.scotusblog.com/case-files/cases/maryland-v-king/`.

The total effect of a policy on crime could depend on either or both of these effects: Criminals can be deterred from offending when they are free, and physically prevented from offending when they are in prison.

For several decades the U.S. has focused on increasing s, sentencing more and more people to longer and longer prison terms. However, incarceration is expensive and, after much study, appears to have limited benefits in the form of a deterrent effect, perhaps due to offenders' high discount rates (Chalfin and McCrary, 2014). As a result, there is increasing interest in the other determinant of the expected cost of crime: the probability of getting caught, p. Whether policies aimed at increasing p will prove effective depends largely on the degree of rationality of criminal offenders. Whether offenders respond rationally to incentives is an open question in the literature (Hansen, 2015), given high rates of mental illness and substance abuse in this population. Evaluating the impact of DNA databases and similar tools should help us better understand the determinants of criminal behavior, and, therefore, how to cost-effectively improve public safety.

There is a small but growing literature on how offenders respond to the probability of punishment. A low-tech way to increase p is to increase the size of the police force, and there is evidence that larger police forces decrease crime rates.[10] However, police officers' effects are typically local and difficult to scale. Another low-tech intervention, sunlight, reduces robbery rates by increasing visibility (Doleac and Sanders, 2015). However, ambient light is difficult to control in many settings. More promising is a program called 24/7 Sobriety that randomly calls probationers in for drug and alcohol testing, with short but immediate jail sentences if they fail. The latest evidence suggests it is extremely effective at reducing DUIs and other substance-related offenses (Kilmer et al., 2013).

Most high-tech innovations — like DNA databases — work mainly by increasing p for reoffenders. These innovations also exhibit increasing returns to scale and can easily target subsets of the population. Examples include GPS monitoring, which has reduced recidivism among released convicts in Argentina (DiTella and Schargrodsky, 2013), and body cameras, which have reduced the use of excessive force by police officers, and/or false complaints by citizens (Ariel et al., 2014). However, not all such interventions have their intended effects. For instance, sex offender registries aim to decrease crime by increasing surveillance of known sex offenders. Agan (2011), Prescott and Rockoff (2011), and Carr (2014) have shown that these registries do not reduce recidivism among registered sex offenders, possibly because the public nature of the registries creates a stigma that makes it difficult for ex-convicts to reintegrate into society.

Many more such programs have yet to be evaluated. Gunshot sensors increase the likelihood and speed with which police officers can arrive at the scene of gunfire; this should increase the apprehension of violent offenders, but there is mixed anecdotal evidence and

[10] See Levitt (2004), Evans and Owens (2007), and Chalfin and McCrary (2014).

no rigorous empirical evidence regarding their impact.[11] Similarly, we know little about the effects of surveillance cameras, CompStat, NIBIN, or other such tools that have become staples of modern police departments.

So far there has been very little research on the effects of DNA databases. Roman et al. (2008) conducted a field experiment in five communities to test the cost-effectiveness of collecting DNA evidence in high-volume property crimes like burglary. The authors found that investigators identified more suspects and had more cases accepted for prosecution when they analyzed and uploaded DNA evidence from these crime scenes. Note that this study speaks more to the potential of DNA databases than how they were actually used in practice. That is, researchers were collecting and analyzing DNA evidence in contexts where this was not standard practice; DNA evidence from property crime scenes has not been a priority for most police departments until very recently. Bhati (2010) proposed and tested a structural model of recidivism using data on criminal histories and DNA profiling from Florida; under relatively strict assumptions, he found 2-3% reductions in recidivism risk attributable to DNA profiling for robbery and burglary, but increases in recidivism risk for other categories. The model's primary identification assumption is that the deterrent and probative effects of DNA profiling are separable, but any deterrent effect is crucially dependent on an expected probative effect (that is, the increased probability of getting caught). In addition, it is likely that the deterrent effect begins at the time of DNA sample collection, which is quite salient, not on the (known to the researcher, but unknown to the offender) date that the profile is ultimately uploaded to the database.

This paper uses criminal history data from a wider sample of states, new data on expansion timing and database size, and exogenous variation in DNA requirement, to investigate both the individual-level and aggregate effects of DNA databases as implemented in the United States through 2010.

3 Effects on Individual Offenders

3.1 Data

In the individual-level analysis, I test the causal effect of the DNA requirement on recidivism, using the effective dates of database expansions as thresholds in a fuzzy regression discontinuity (RD) framework. To do this, I need detailed criminal histories for individual offenders and information on their states' database expansions.

I obtained longitudinal criminal history data from the Departments of Correction (DOCs) in seven states: Florida, Georgia, Missouri, Montana, New York, North Carolina, and Penn-

[11] Carr and Doleac (2015) describe ShotSpotter data and consider their value for studying policy effects on gun violence. Like other surveillance tools, part of ShotSpotter's potential benefit is collecting higher-quality data on criminal behavior.

sylvania.[12] These are detailed, offense-level data that include unique identifiers for individual offenders, as well as the associated dates of incarceration (including conviction and release dates), offense type(s), and offenders' birthdate or current age, sex, and race. Due to their very limited number, I exclude female offenders from my analysis.

These data do not include information on actual DNA collection or profiling. Instead, I use the details of each offender's criminal history to determine upon which release date (if ever) a DNA sample was *required*, given his state's DNA database laws. This necessitated matching offenses as coded in DOC datasets to those listed in DNA database statutes. To ease comparison across states, I also determine which offenses count as FBI Index I crimes, which have standardized definitions.

Information on the timing and details of state DNA database expansions comes directly from state legislative histories.

The RD design requires identifying discontinuities in the probability of DNA requirement. Appendix Figures A-1 through A-8 plot the weekly probabilities that released offenders were required to provide DNA over time, by state, for each serious offense type. The dates of database expansions are shown by red vertical lines. Because released offenders were often serving time for multiple convictions, the probabilities do not typically jump from 0 to 1 at the time the relevant qualifying offense was added. For instance, consider an inmate who is serving time for both burglary and larceny. If burglary (but not larceny) is a qualifying offense, the inmate will be required to provide a DNA sample upon release, and he might show up as "treated" in both the post-expansion burglary and pre-expansion larceny samples. This type of crossover between offense categories explains why there are sometimes two or more discontinuities in these graphs, and why a fuzzy RD design is appropriate.[13]

To maximize statistical power, I group (separately) the serious violent and serious property offenders for my main analyses.[14] Figures 1 and 2 plot, by state, the probabilities that released offenders convicted of any serious violent or any serious property offense are required to provide DNA upon release. Because several offense types are grouped together, it is now even more likely that more than one database expansion will cause a discontinuous increase in DNA requirement for offenders in each state's sample.

To create offender samples for the empirical analysis, I take the individuals incarcerated for the crime category of interest, released within one year (365 days) before or after each relevant discontinuity in Figures 1 and 2. Again, this could result in multiple periods of

[12]These states were selected based on availability of data, not due to their specific characteristics. The diverse nature of this sample should limit concerns about external validity, but of course states very different from those in this sample might have had different experiences.

[13]Some of the "fuzziness" is also due to mismatches between the local and FBI definitions of an offense. The most striking disparity in the graphs is for murder, where the FBI definition includes non-negligent manslaughter, but many states included only homicide convicts in their databases at first. Thus, the murder convict sample includes manslaughter convicts whose DNA was not collected after the homicide expansion.

[14]Serious violent offenses are murder, forcible rape, aggravated assault, and robbery. Serious property offenses are burglary, larceny, vehicle theft, and arson.

interest per state for a particular crime (for example, there are two discontinuities and therefore two periods included for serious property offenses in Pennsylvania). I restrict attention to offenders released before January 1, 2007, to ensure that everyone in the sample is observed for a full five years after release. (This necessitates excluding the 2006 expansion in Florida, which added property offenders.) Finally, I recenter the release dates so that 0 is the expansion date, and combine the samples from all relevant periods into one larger sample of offenders.

The final samples include database expansions spanning the years 1994 to 2005. Summary statistics for the serious violent and serious property offender samples are shown in Tables 1 and 2. Figure 3 plots the probability of DNA requirement for these samples. There are clearly large, discontinuous increases in the probability of DNA requirement at the database expansion date for each aggregated sample.

More information on the data, including the dates of the discontinuities and summary statistics for each state dataset, is in Appendix B.

3.2 Empirical Strategy

3.2.1 Checking validity of the RD design

The purpose of the RD design is to divide offenders into treatment and control groups with similar pre-treatment propensities to reoffend. I test that the pre- and post-expansion samples are comparable in two ways: (1) I predict individuals' propensities to reoffend within 3 and 5 years, based on their observable characteristics, then test for discontinuities in this predicted variable. (2) I conduct McCrary tests for discontinuities in the distribution of offenders released around the threshold.

Tables 1 and 2 show the observable characteristics of offenders released before and after the expansion dates, along with the results of t-tests for differences in means. The predicted probabilities of recidivism, based on those observable characteristics, are at the bottom of each table.

Consider the serious violent offenders first: Table 1 shows some differences in individual characteristics between the pre- and post-expansion samples. Convicts released after the expansion date have been incarcerated more times, are less likely to have been convicted of sexual assault, and are more likely to have been convicted of robbery or vehicle theft. There are also differences in the shares of convicts coming from each database expansion sample (more from Florida 1995 and Pennsylvania 1996; fewer from Florida 2002, New York 1999, North Carolina 1994, and Pennsylvania 2005). However, all of these differences are quite small in magnitude. When I predict individuals' probability of reoffending, I find that individuals released post-expansion are not significantly different from those released pre-expansion. If anything, those released after the expansion are slightly *more* likely to reoffend – which would bias estimates upward – but the magnitude of that potential bias is extremely small (less than 1% of the baseline).

Table 2 reveals even smaller differences between pre- and post-expansion samples for the serious property offenders. Convicts released after the database expansions are more likely to be white (less likely to be black), slightly older, and less likely to have been convicted of larceny. There are again some differences in the shares of the aggregate sample coming from individual database expansions (more from Florida 2000; fewer from New York 1999). However, when I predict individuals' probability of reoffending based on their observable characteristics, I find no difference between the pre- and post-expansion samples. In both cases, then, the differences in individual characteristics aren't large or important enough to affect convicts' propensity to reoffend.

Appendix Figure A-9 plots the predicted probabilities of reoffending for each sample. It shows the same effects as in the tables: a very small uptick in the predicted probability of reoffending for serious violent offenders released after the expansion, and no change for serious property offenders. While these results suggest the pre- and post-expansion samples are extremely similar, and that the RD will be unbiased, I will control for offender characteristics in my analyses to ensure that the small differences that do exist are not affecting the estimates.

Next, I consider the possibility that offenders' release dates were intentionally timed to occur before or after the database expansions, potentially introducing selection bias. While any such manipulation of release dates (which were determined at the time of convicts' sentencing, and unaffected by parole decisions) would have been illegal, I check this formally using a McCrary test of the distribution of released offenders around the expansion date. The resulting graphs are presented in Figure 4. There is no evidence of a discontinuity in the distribution of serious violent offenders: the distribution appears smooth and any difference is statistically insignificant. However, there does appear to be an increase in the number of serious property offenders released just after the expansion date. This discontinuity is statistically significant ($t = 3.29$). While this discontinuity in releases does not necessarily mean the estimates based on these data will be biased, it does call for additional scrutiny.

This aggregated sample of serious property offenders is the combination of data from eight separate database expansions, which allows me to construct more robust subsamples. I conduct separate McCrary tests using the data from each individual expansion in the serious property offender sample; those graphs are presented in Appendix Figure A-10. Three expansions appear particularly problematic based on this validity test: Florida 2000, Missouri 2005, and Pennsylvania 2002. When I drop those expansions, the resulting McCrary test for the remaining sample reveals no statistically significant discontinuity; the resulting graph is in Appendix Figure A-11. Alternatively, I could use only the expansions that show completely smooth distributions through the expansion date threshold. The graphs of two expansions look ideal in this way: Georgia 2000 and North Carolina 2003. Using data from only those expansions produces a smooth aggregate distribution, also shown in Appendix Figure A-11. I will use each of these two constructed subsamples to check the robustness of the full-sample serious property offender results. If the full-sample discontinuity is a statistical anomaly – instead of evidence of selection – these robustness samples should produce similar results.

3.2.2 Estimating the effects of DNA requirement on recidivism

I first test the reduced form effect of the database expansions, using *Post* as the indicator of treatment (being released post-expansion). I estimate the following linear probability model using OLS:

$$\begin{aligned}\mathrm{Pr(Recidivate)}_j &= \beta_1 + \beta_2\,\mathrm{Post}_j + \beta_3\,\mathrm{f(Release\ Date)}_j \\ &\quad + \beta_4\,\mathrm{Post{*}f(Release\ Date)}_j + \lambda_{state*year} + \gamma_{month} + \beta_5\,\mathrm{X}_j + \mathrm{u}_j, \quad (2)\end{aligned}$$

where j indexes individual offenders. $\lambda_{state*year}$ are interactions of state and year fixed effects, so the identification comes from within-state-year variation in the probability that DNA is required. γ_{month} are month-of-year fixed effects, which control for any effect that month of release has on recidivism risk. X_j is a vector of demographic and criminal history information, including: race, age, types of previous convictions, and whether this was the individual's first incarceration. These controls should absorb a large share of the variation in recidivism risk, and allow me to more precisely detect the effect of the treatment. However, they should have little effect on the estimates since the pre- and post-expansion samples are balanced.

The advantages of this reduced form approach are transparency and ease of interpretation. However, it does not utilize the information about the intensity of the treatment. This is important here because database expansions did not typically increase the probability of DNA requirement from 0 to 1, so the reduced form approach will underestimate the effect of the treatment on individuals' behavior.

To account for this information, I use the following fuzzy RD specification, estimated using an IV 2SLS model, which effectively scales up the previous estimates based on the proportion of individuals affected by the expansion:

$$\begin{aligned}\mathrm{Pr(Recidivate)}_j &= \beta_1 + \beta_2\,\mathrm{Pr(DNA\ required)}_j + \beta_3\,\mathrm{f(Release\ Date)}_j \\ &\quad + \beta_4\,\mathrm{Pr(DNA\ required)\,{*}f(Release\ Date)}_j + \lambda_{state*year} + \gamma_{month} + \beta_4\,\mathrm{X}_j + \mathrm{w}_j, \quad (3)\end{aligned}$$

where *Pr(DNA required)* is estimated in a first stage regression, using *Post* as an instrument:

$$\begin{aligned}\text{Pr(DNA required)}_j = \delta_1 + \delta_2\,\text{Post}_j + \delta_3\,\text{f(Release Date)}_j \\ + \delta_4\,\text{Post*f(Release Date)}_j + \lambda_{state*year} + \gamma_{month} + \delta_4\,\text{X}_j + \text{v}_j \end{aligned} \quad (4)$$

For the sake of consistency across samples, my preferred bandwidth is 365 days (one year) before and after the release date, for both the serious violent and serious property samples. (Imbens-Kalyanaraman optimal bandwidth estimates are slightly larger, and differ across samples. Appendix Figure A-12 shows each of these bandwidths.) In my preferred specification, f(ReleaseDate) is a linear function. I will verify that the results are not sensitive to these choices in Section 3.4.

The outcome variables are conviction of any offense within three and five years of release. In both the reduced form and RD specifications, β_2 is the coefficient of interest.

Figure 5 plots the residuals of the predicted outcome regressions estimated in Section 3.2.1 (residual = observed recidivism - predicted recidivism). It appears the DNA requirement decreases the probability of reoffending for both samples.[15] This previews the effect I find using the reduced form specification.

3.2.3 Interpretation of β_2

There are three points to keep in mind when interpreting the effect of the DNA requirement on recidivism:

First, as described in the previous section, I do not observe actual DNA collection, but determine the requirement to submit a DNA sample based on individuals' criminal histories and incarceration spells. The coefficient β_2 therefore measures the effect of the intention to treat with DNA profiling. This might not be the same as the effect of DNA profiling if states were slow to implement the new laws and/or if some offenders were mistakenly released before providing a DNA sample.[16] To the extent that this occurred, β_2 will be biased toward zero, particularly for earlier expansions and when using smaller bandwidths.

Second, measuring recidivism accurately is difficult because offenses are only observed if the offender gets caught. That is, instead of the ideal outcome variable, Pr(Reoffend), I

[15] Appendix Figure A-13 plots the raw outcome data (probability of a new conviction within 5 years) against the release date; we see a similar decrease for both samples.

[16] Statistics on the frequency of such mistakes are unavailable for the states in my sample during the time period of interest, and it is difficult to estimate how large an effect this might have on my results. Even in more recent years, despite much more experience, states have a difficult time implementing their policies perfectly: New York State requires DNA collection upon intake to the Corrections system, and estimated a 92% collection rate within 2 months of an eligible sentence to a jail or prison in 2009. [http://criminaljustice.state.ny.us/pio/annualreport/2009-crimestat-report.pdf, page 20]

observe Pr(Reoffend and Convicted), where

$$\text{Pr(Reoffend and Convicted)} = \text{Pr(Reoffend)} * \text{Pr(Convicted} \mid \text{Reoffend)}\,. \tag{5}$$

DNA profiling should affect both factors on the right-hand side in equation 5, in opposite directions. If DNA profiling helps law enforcement identify a crime's perpetrator, as designed, it increases Pr(Convicted | Reoffend); this is the probative effect. This increases E(Cost) in equation 1 and should therefore reduce Pr(Reoffend); this is the deterrent effect. The coefficient β_2 estimates the *net* effect of a DNA requirement. My data and identification strategy will not allow me to separate these two effects. Because the probative and deterrent effects cancel each other out to some extent, a significant positive estimate should be interpreted as a lower bound on the true probative effect, and a significant negative effect should be interpreted as a lower bound on the true deterrent effect. (Note that a zero net effect could mean that DNA databases have no effect on offenders, or that the deterrent and probative effects cancel each other out completely.)

The third point regarding β_2 has to do with generalizability. This identification strategy depends on testing the effects of DNA databases immediately after they were expanded. It is quite possible, even likely, that the effectiveness of this law enforcement tool grows over time as police learn how to use it and offenders learn (perhaps through personal experience) of its probative effect. It is also possible that offenders gradually learn how to avoid detection by DNA analysis. Therefore, it is not clear whether the short-term effects found in this part of the paper should be thought of as upper or lower bounds on the longer-term effects. For this reason, the aggregate effects of DNA databases on crime over the longer term—presented in Section 4—will be an important supplement to this analysis.

3.3 Results

Table 3 presents results of the first-stage regression for the 2SLS analysis. Being released post-expansion clearly has a large, positive effect on the probability of DNA requirement: 0.42 for serious violent offenders and 0.78 for serious property offenders (on a 0–1 scale). Both estimates are highly statistically significant ($p < 0.01$).

Table 4 shows the estimated effect of the DNA requirement on serious violent offenders in the first panel, and serious property offenders in the second panel. Coefficients show the percentage point change in the probability of conviction for any offense within three years (Columns 1–6) and five years (Columns 7–12). Many studies use three-year horizons to measure recidivism, but it is clear in this case that it takes five years for statistically significant effects to become evident, perhaps because it takes a while for offenders to learn of DNA's potential and so the deterrent effect is delayed. I present results for both outcome measures, though my preferred specification uses the five-year horizon.

Column 1 shows the reduced form effects of DNA requirement on three-year recidivism, controlling only for the running variable and state-by-year fixed effects. Being released post-expansion reduces the probability of recidivating within three years by 1.3 percentage points (7%) for violent offenders; this effect is marginally significant ($p < 0.10$). The effect on property offenders is negative but small and statistically insignificant.

Column 2 adds month-of-year fixed effects, to control for seasonal effects of the release date on offenders' outcomes. Adding this control has essentially no effect on the estimate for violent offenders. The estimate for property offenders is now positive, but still small and statistically insignificant.

Column 3 adds a variety of demographic and criminal history variables. The results are nearly identical when these additional controls are added, due to the similarity of the pre- and post-expansion samples.

Columns 4–6 show the equivalent results using an IV 2SLS model to estimate the fuzzy RD specification. These results incorporate information about the intensity of the treatment (that is, the increase in the probability of DNA requirement at the threshold), so are generally larger than the reduced form estimates. Over the three-year time horizon, the impact on recidivism is marginally significant only for serious violent offenders. For that group, the DNA requirement results in a 2.9 percentage point (16%) decrease in recidivism, and that estimate changes very little when additional controls are added. The effect on serious property offenders remains small and statistically insignificant across the three specifications.

Columns 7–12 repeat these analyses, using five-year recidivism as the outcome. Over this longer time horizon there is more variation in the outcome measure, and the effect of the DNA requirement becomes clearer.

Column 7 shows the reduced form effects, controlling only for the running variable and state-by-year fixed effects. Being released post-expansion reduces the probability of recidivating within five years by 1.9 percentage points (7%) for serious violent offenders; this effect is statistically significant at conventional levels ($p < 0.05$). For serious property offenders, the estimate is a 1.8 percentage point (5%) reduction; this effect is marginally significant ($p < 0.10$).

Column 8 adds month-of-year fixed effects. This increases the estimated effect of being released post-expansion, for violent offenders: the coefficient is now -2.4 percentage points (9%), and is highly statistically significant ($p < 0.01$). However, adding this control reduces the estimated effect for property offenders: the coefficient is now about half the size and statistically insignificant.

Column 9 adds demographic and criminal history controls; the results are nearly identical to those in Column 8.

Column 10 shows fuzzy RD results, controlling for the running variable and state-by-year fixed effects. Based on this specification, DNA requirement results in a 4.5 percentage point (17%) reduction in recidivism for serious violent offenders over five years; this effect is statistically significant ($p < 0.05$). For serious property offenders, the impact is a marginally-significant 2.4 percentage point (6%) reduction in recidivism.

Column 11 adds month-of-year fixed effects, with similar effects as before. In this specification, the effect of DNA requirement on 5-year recidivism is -5.5 percentage points (21%) for serious violent offenders; this estimate is highly statistically significant ($p < 0.01$). For property offenders, DNA requirement reduces 5-year recidivism by 1.2 percentage points (3%), but this effect is statistically insignificant.

Column 12 adds demographic and criminal history controls, which changes the effect magnitudes only slightly from those in Column 11. Based on this specification, DNA requirement results in a 5.7 percentage point (21%) reduction in recidivism for serious violent offenders over five years; this effect is highly statistically significant ($p < 0.01$). For serious property offenders, the impact is a statistically insignificant 1.2 percentage points (3%) reduction in recidivism.

Tables 5 and 6 present results separately for offenders with specific criminal histories. These analyses suffer from limited sample sizes, but provide useful information about which subgroups of offenders are most affected by the DNA requirement, over the 3- and 5-year periods. It appears that effects on rape, robbery, and larceny convicts are contributing most heavily to the aggregate results: the effects on 5-year recidivism are -32%, -20%, and -20%, respectively, for these groups. These groups also appear to be driving the difference between the 3- and 5-year estimates. In addition, we can see for which groups adding the full set of controls affects the estimates. For rape, assault, robbery, larceny, and vehicle theft convicts, adding the full set of controls tends to increase the magnitude of estimates. For burglary convicts, adding those controls reduces the size of the effect – and this appears to be driving the reduction in the effect size for the full property sample when controls are added.

3.4 Robustness checks

I conduct a variety of additional tests to consider the robustness of the above results:

Recall that the full sample of property offenders failed the McCrary test, so I constructed two subsamples with smooth distributions through the threshold. Appendix Table A-1 presents estimated effects, using those two robustness samples of property offenders: (1) dropping Florida 2000, Missouri 2005, and Pennsylvania 2002; and (2) only including Georgia 2000 and North Carolina 2003. The sample sizes are now much smaller, so the estimates are less precise, but the coefficients in both cases are, if anything, larger in magnitude than those in panel 2 of Table 4. This implies that the above results for property offenders are not being biased by the discontinuous increase in releases post-expansion.

Appendix Table A-2 further considers the extent to which the main results are being driven by particular database expansions. Each column drops a different expansion, in turn, from the analysis. Results are quite consistent across the different samples for both violent and property offenders.

The choice of bandwidth can matter a great deal in RD specifications, so I want to be sure that the 365-day bandwidth used above produces estimates that are consistent with those from other bandwidth options. Appendix Figure A-12 plots estimated coefficients for the preferred specification (Column 10 in Table 4), using bandwidths ranging from 50 to 700 days. The estimates reported above are within the confidence intervals of almost all the other other estimates, and are quite similar to the estimates for bandwidths larger than approximately 225 days. In particular, note that the optimal bandwidths calculated using the Imbens-Kalyanaraman (IK) method, as well as the IK method for fuzzy RDs, are larger than the 365-day bandwidth used in my preferred specification. The latter was chosen for consistency across samples, but for transparency I have included vertical lines in the coefficient dot-plots denoting the IK optimal bandwidth sizes, along with the 365-day bandwidth. The coefficient-sizes at those optimal bandwidths are nearly identical to those found when using a 365-day bandwidth.

For very small bandwidths, the estimates are different but the confidence intervals are quite wide. For violent offenders, estimates hover near-zero until the 200-day mark. This could be due to delays in implementing the DNA database expansion; any such delays would lead me to incorrectly label some post-expansion offenders as treated when they were not. For property offenders, the coefficients are large and negative for small bandwidths but shrink as the bandwidth widens, leveling off at the effect-size reported above.

Finally, some might worry that the assumption that the running variable has a linear effect on recidivism is too strict. Appendix Table A-3 presents results as the functional form of the running variable changes. While higher-order polynomials quickly strain the statistical power of these data, all estimates are well within the confidence intervals of the others.

4 Effects on Crime Rates

The effect on profiled offenders is not the only determinant of DNA databases' effects on crime. Factors such as the reactions of unprofiled and never-offenders, the use of forensic evidence by law enforcement, the response to such evidence from jurors and judges[17], and changes in the behavior of crime victims and the general public all contribute to the net effect of this law enforcement tool.

To address the most policy-relevant question, I investigate the causal effect of DNA

[17] In particular, there is anecdotal evidence that jurors have come to expect high-quality forensic evidence in all types of cases. This is commonly referred to as the "*CSI* effect"; see Owens (2010) for a review of the available evidence on the existence of this effect.

database size on crime rates. DNA databases are designed to affect crime by increasing the probability that a known offender gets caught if he reoffends. Thus, the effectiveness of a state's DNA database increases with the probability that a potential offender is in the database; this probability is the intensity of treatment. In the years soon after each database expansion, this probability is highly correlated with the size of the database relative to the state population.[18]

The problem with this treatment variable is that database size is endogenous. The number of criminal offenders and offender DNA profiles in a state are simultaneously determined and positively correlated, so OLS estimates of the effect of database size on crime rates will be biased upwards. At the same time, states' adeptness and motivation with regard to implementing database laws will also affect database size. Because these state characteristics might affect crime rates through other channels, OLS estimates could suffer from omitted variable bias. I therefore need an instrument for database size.

There is substantial anecdotal evidence that database expansions were frequently triggered by "if only" cases – high-profile murders or rapes that could have been prevented "if only" the offender had been added to the database after an earlier, more minor conviction or arrest. Such incidents are typically unrelated to underlying crime trends, and so the timing of database expansions should be exogenous shocks to state database size. I will first confirm that the variation in law timing is exogenous, then use that variation to construct instrumental variables for database size. Finally, I'll use these instruments to test the effect of database size on state crime rates.

4.1 Data

I consider the effect of DNA databases during the years 2000-2010. This is relatively early in their development, so database size should be highly correlated with the intensity of treatment.[19]

To ease comparison across states, I group database expansion types into four categories: felony sex offenses, felony violent offenses, felony burglary, and all other felony offenses. Legislated expansions can apply to new convicts and/or anyone incarcerated on or after the effective date. Information on database expansion timing comes directly from state legislative histories. The year of each category of expansion, by state, is shown in Appendix Table A-4.

[18] This is because most profiles are from offenders who are still active. However, as those profiled offenders get older, die, or move out of state, their profiles are not deleted, so databases continue to grow. The number of profiles per capita in the database therefore becomes a less useful proxy over time for the probability than an active offender will get caught.

[19] I focus on years 2000 and later because a change in DNA analysis technology dramatically decreased the cost of analyzing DNA samples in that year. It also required states to reanalyze any samples they had previously uploaded to their databases. In this sense, 2000 served as a "reset" year, and DNA database expansions should affect crime rates differently before and after this point.

Historical data on state DNA database size did not previously exist, and so I constructed a dataset using a variety of sources – primarily data from state forensic agencies and media reports. (See Appendix B for more information.) The resulting annual data on database size are likely measured with error, but instrumenting for them will remove any attenuation bias. Note that data on overall database size are the best available; more information (such as the type of offender each profile came from, or the offender's age) would be very helpful, but current database structures don't allow those types of queries.[20]

Annual data on crime rates come from the FBI's Uniform Crime Reports (UCR). I use the following offense types: murder, forcible rape, aggravated assault, robbery, burglary, larceny, and vehicle theft. (I exclude arson because there are too few arsons each year to produce meaningful estimates.) I consider effects on those crimes individually, and aggregated to violent crimes (murder, forcible rape, aggravated assault, robbery) and property crimes (burglary, larceny, and vehicle theft).

I create a simulated instrument by predicting the number of qualifying offenders in each state, for each year. These predictions are based on pre-period rates of newly-convicted offenders and incarcerated inmates, as well as the timing of database expansions.

Alternative IV specifications use estimates of pre-period reported offenses, arrests, convictions, or releasees as proxies for newly-convicted offenders. Data on pre-period reported offenses comes from the 1999 UCR. Arrests are based on the number of cleared offenses from the 1999 UCR. Convictions are based on the number of cleared offenses, multiplied by the share of arrested offenders who were convicted, by type of crime (using 2000 data from the State Court Processing Statistics). Releasees are based on convictions, lagged by the expected time served for each crime type (from the 1999 National Corrections Reporting Program). I also estimate the share of arrests, convictions, and releases of offenders under age 40 (when offenders are more likely to be active), based on arrestee and convict demographics from the 2000 State Court Processing Statistics. My preferred specification uses the number of pre-period convictions as a proxy for newly-convicted offenders, since it is closest in concept to the actual number of interest, but note that the results are very similar across all of these IV options.

Incarcerated inmates are defined as the number of offenders incarcerated for a particular offense. Data on pre-period prison inmate populations by state come from the 1999 National Jail Census, maintained by the Bureau of Justice Statistics (BJS). I estimate the share incarcerated for each type of offense based on national data, also from the BJS.

For the years 2000-2010, these data yield an unbalanced panel of 344 state-by-year observations where both database size and simulated instruments are available. In all cases the rates are determined by dividing by the state population, from the U.S. Census. Summary statistics are shown in Table 7.

[20]Based on conversations with several state database administrators.

Finally, in a robustness check I consider the effect of DNA database size on the probability of making an arrest in new crimes. Data for the probability of making an arrest come from the FBI's National Incident-Based Reporting System (NIBRS), for 2000-2010. NIBRS is not available for all states in all years, but provides much richer data on each reported crime in a jurisdiction — including, most crucially, whether an arrest was made in a particular case. These data yield 1.9 million violent offenses and 15.3 million property offenses in states and years where database size and simulated instruments were available.

4.2 Empirical Strategy

4.2.1 Exogeneity of database expansion timing

Within-state variation in database size comes largely from legislation adding new types of offenders. Before using these shocks to database size to construct instrumental variables, I want to make sure that they are exogenous with respect to other factors that might affect crime trends.

Appendix Tables A-5 through A-8 show the effects of state characteristics (including racial composition, poverty rate, educational attainment, political leanings, and pre-period violent crime rates) on the timing of database expansions, by type of expansion. The top panel in each table shows the effects of state characteristics on the year of each expansion; the bottom panel shows the effects of state characteristics on whether a state's expansion was relatively late (after the median year). Very few estimates are statistically significant, and their signs are inconsistent, showing no clear patterns.

Figure 6 plots violent and property crime rates by year, relative to the timing of what was typically the first substantive database expansion in each state: adding new violent crime convicts. This is the cleanest event, as preceding years are less likely to be confounded by other expansions. While both violent and property crime rates were increasing in pre-expansion years, we see a clear break in trend at year 0, when the databases were expanded to add violent crime convicts. Appendix Figures A-14 through A-18 show similar plots for other types of database expansions. These are more difficult to interpret because of the cumulative nature of these law changes. We see effects at year 0 in the cases of adding burglary convicts, violent inmates, burglary inmates, and (to a lesser degree) adding all felony inmates. In other graphs (for instance, for the addition of all felony convicts), we see the trend change a couple years before the expansion of interest, but again, those years are confounded by the (perhaps larger) effects of previous database expansions – for instance, violent and burglary convicts.

Overall, these provide strong evidence that the dates of database expansions were independent of pre-existing trends and characteristics that might affect future crime rates. I will therefore use the timing of these expansions to construct instrumental variables.

4.2.2 Constructing simulated instrumental variables

I predict the number of qualifying offenders using the following equation:

$$\text{Predicted Qualifying Offenders}_{s,t} = \sum_j [\text{1999 Conviction Rate}_{s,j} * \text{Years Convicts Included}_{s,t,j} + \sum_j [\text{1999 Inmate Rate}_{s,j} * \text{I(Inmates Included}_{s,t,j})], \quad (6)$$

where s indexes states, t indexes years, and j indexes offense categories. In words, the predicted number of qualifying offenders (per 100,000 residents) is a function of what the flow of new qualifying offenders *would have been* if conviction rates remained at 1999 levels, multiplied by the number of years new convicts were included; plus what the stock of qualifying inmates *would have been* if prison populations remained at 1999 levels, multiplied by an indicator of whether inmates were included. For instance, consider a state that adds convicted and incarcerated burglars to its database in 2002, and in 1999 had 50 burglary convictions per 100,000 residents and 200 incarcerated burglars per 100,000 residents. The simulated instrument would be 0 in 2001, 250 in 2002 (50 convictions * 1 year + 200 inmates), 300 in 2003 (50 convictions * 2 years + 200 inmates); 350 in 2004 (50 convictions * 3 years + 200 inmates), and so on. However, the predicted effect on qualifying offenders would be zero for a state that had no burglary convictions or incarcerated burglars in 1999 (0 convictions * n years + 0 inmates). I want to distinguish between these two states: the first should see a larger effect on crime than the second because the database expansion was more meaningful. The instrument makes this distinction by quantifying the intensity of the treatment.

This resulting variable, *predicted qualifying offenders*, will be the instrument for database size. It quantifies the effects of the law changes by "simulating" the number of offenders who should be included in the database in each year, based on legislated qualifying offenses and pre-period (1999) conviction rates and prison populations. By estimating both the stock and flow of qualifying offenders, I produce an instrument that is strongly correlated with the actual number of profiles (F-statistic = 29), but uncorrelated with crime rates through any other channel. Using pre-period statistics eliminates the simultaneity problem, and using the number of qualifying offenders — rather than the number of uploaded profiles — corrects the omitted variable bias. This IV approach therefore eliminates the biases that affect the OLS estimates. Subject to the identifying assumption that legislation timing does not depend on pre-existing crime trends, *predicted qualifying offenders* is a valid instrument for the actual number of profiles in a state's database.

As described above, I create several alternative IVs as robustness checks. Instead of using 1999 convictions to predict the flow of newly-convicted offenders, they use: (1) reported offenses, (2) arrests, or (3) releases. Second versions of the arrest, conviction, and release

IVs estimate the share of new offenders who were under age 40 (and thus more likely to reoffend) in each case. We can think of these under-40 IVs as estimating the number of likely reoffenders in the database, though of course the data on actual database size cannot be broken down by offenders' ages.

4.2.3 Estimating the effect of database size on crime rates

To estimate the causal effect of DNA databases on crime rates, I run a 2SLS instrumental variable regression of crime rates on database size, relative to the state population:

$$\text{Crime Rate}_{s,t} = \beta_1 + \beta_2 * \text{Profile Rate}_{s,t} + \beta_3 * \text{Police Per Capita}_{s,t} + \lambda_{region*year} + \gamma_{state} + w_{s,t}, \quad (7)$$

where s indexes states, and t indexes years. Including state-by-year fixed effects is not possible here, because that is the level at which the treatment varies. However, I do include state fixed effects and region-by-year fixed effects, based on U.S. Census Regions; the latter control non-parametrically for variation in crime rates over time, at the regional level. The specification also includes the number of local police officers per capita, to proxy for other state-level crime-reduction strategies. β_2 is the coefficient of interest.

Profile rate is estimated by the following first stage regression:

$$\text{Profile Rate}_{s,t} = \delta_1 + \delta_2 * \text{Predicted Qualifying Offenders}_{s,t} + \delta_3 * \text{Police Per Capita}_{s,t} + \lambda_{region*year} + \gamma_{state} + v_{s,t}, \quad (8)$$

where s indexes states, and t indexes years. *Predicted qualifying offenders* is the instrument from equation 6.

4.3 Results

4.3.1 Effects on Crime Rates

First stage results are shown in Table 8 for my preferred IV, which uses convictions to estimate the number of qualifying offenders. Appendix Table A-9 shows first stage results for each of the alternative IVs described above. F-statistics range from 22.3, when using reporting offenses as a proxy for newly-convicted offenders, to 50.9, when using releasees under age 40 as a proxy for newly-convicted offenders. All are strong instruments by conventional standards (F-statistics > 10). Since readers might prefer instruments other than my preferred IV, I will present my main results using each of these alternatives.

The estimated effects of database size on crime rates are presented in Table 9. The top two panels show effects on violent crime and property crime, respectively. The remaining panels show results by individual crime type.

Column 1 shows OLS estimates of equation 7, without the instrumental variable. As expected, these are biased upwards. The OLS specification does not estimate a statistically-significant effect of DNA databases on crime, in any case. For the burglary rate, the coefficient is positive, the opposite of what we would expect.

Columns 2–8 show IV estimates using *predicted qualifying offenders* as an instrument for database size.

Results from my preferred IV specification are shown in Column 5. Here we see negative and statistically significant estimates for all crimes except burglary. The coefficients suggest that one additional DNA profile (per 100,000 residents) results in 0.051 fewer violent crimes and 0.323 fewer property crimes (per 100,000 residents). These estimates are statistically significant ($p < 0.01$ and $p < 0.05$, respectively) but imprecise. The 95% confidence interval for violent crime ranges from -0.014 to -0.088 offenses. The average database grew by 2183 profiles (per 100,000 residents) over this time period, so this confidence interval implies that DNA databases reduced annual violent crime rates by 7–45%, between 2000 and 2010, relative to the counterfactual. For property crimes, the confidence interval ranges from -0.075 to -0.571. This implies that, on average, DNA databases reduced annual property crime rates by 5–35% between 2000 and 2010.

The remaining panels consider the effects on individual crime types. An additional DNA profile reduces murder by 0.0004 offenses (relative to a baseline of 4.776), rape by 0.0045 offenses (relative to a baseline of 33.981), assault by 0.0347 offenses (relative to a baseline of 280.35), and robbery by 0.0115 offenses (relative to a baseline of 107.96). These consistently-significant effects on violent crime might be expected, given the focus of DNA evidence collection on violent crime scenes during this period. We would expect a greater deterrent effect on violent crime for profiled offenders, if they think DNA evidence is more likely to be available at such scenes. (Indeed, this is what I found in Section 3). In addition, serial violent offenders should be more likely to be caught and quickly incapacitated once their DNA profile is in the database, when jurisdictions prioritized analyzing DNA evidence from violent crime scenes.

However, as the aggregate property crime effects show, we see that DNA databases reduce property offense rates as well. An additional DNA profile reduces larceny by 0.2087 offenses (relative to a baseline of 2496) and vehicle theft by 0.0934 offenses (relative to a baseline of 353.9). DNA databases do not have a statistically-significant effect on burglary, though the coefficient is negative. As Roman et al. (2008) showed, DNA evidence in burglaries is quite useful for identifying offenders, but it was not consistently collected and analyzed during this period. The effects on other property crimes could be the result of non-specialist offenders who are deterred or incapacitated due to a violent crime. Data from the first part of this paper shed some light on this: Appendix Figures A-6 and A-7 show that a substantial number of larceny and vehicle theft convicts were added to the database for other, more serious crimes (these are the non-zero observations, before the expansion that shifts most to 1). This suggests those types of offenders frequently cross over into other types of criminal activity. In contrast, Appendix Figure A-5 shows very little such crossover by burglars. These patterns are suggestive but are consistent with the observed effects on property crime rates.

The remaining columns use alternative constructions of the instrument. Column 2 uses pre-period reported offenses instead of convictions to predict the flow of newly-convicted offenders. Column 3 uses pre-period arrests. Column 4 uses pre-period arrests of offenders under age 40. Column 6 uses pre-period convictions of offenders under age 40. Column 7 uses pre-period releasees. Column 8 uses pre-period releasees under age 40. These alternative instruments vary in their statistical power, but across the board the estimated effects are very similar, in terms of magnitude and statistical significance, to those in Column 5.

4.3.2 Are the benefits of adding marginal offenders non-linear?

The above analysis estimates the average effect of adding a marginal convicted felon to a state's DNA database. Given ongoing policy debates about adding more categories of offenders, such as arrestees and misdemeanor convicts, to state databases, it is important to consider whether the the effect of adding marginal offenders is non-linear.

We might expect increasing returns to marginal profiles if minor offenders are likely to commit serious offenses in the future. That is, if there is crossover between crime types, then adding offenders earlier in their criminal careers could have bigger benefits than adding already-violent offenders who are spending many years in prison and/or are already on law enforcement's radar. (The stated goal of adding minor offenders is typically not to prevent future minor offenses, it's to catch serious violent offenders more quickly.)

To directly test whether the crime-reducing effect of a marginal profile increases or decreases as databases grow, Appendix Table A-10 adds a quadratic term (the square of DNA database size) to the main analysis. For both violent and property crime, the coefficient on the quadratic term is statistically insignificant. This means I cannot reject the hypothesis that the effect of DNA profiles on crime is linear. However, note that the sign on the quadratic term is negative for violent crime and positive for property crime. This implies that marginal profiles have increasing returns when it comes to reducing violent crime, and decreasing returns when it comes to reducing property crime.

4.4 Robustness checks

It is possible that the relationship between the simulated instrument (qualifying offenders) and database size is not linear, and this reduces its power. Appendix Table A-11 allows a more flexible function (adding quadratic and cubic terms) of the instrumental variable. In general, this reduces the F-statistic, and has only a small effect on the estimates.

In addition to the event study graph in Figure 6, some readers might like more rigorous evidence that pre-existing crime trends are not driving the empirical estimates. Appendix Table A-12 shows the effects of adding three leads and three lags of the simulated instrument, as well as a lag of the outcome measure. Note that in this case I am using reduced-form estimates of the IV's direct effect on crime rates, not 2SLS estimates. This is because the unbalanced panel of database size precludes adding leads and lags without losing many observations. The magnitudes of the coefficients will therefore not be comparable to the main estimates, but their signs and significance should be.

We see that the effect of the IV on violent crime is largely robust to the addition of leads and lags. The apparent exception is when using an IV based on releasees. However, note that the standard errors on that estimate are large, so I cannot reject the hypothesis that the coefficient is the same as without the leads and lags. In addition, the releasee IV is essentially a lagged version of the convict IV, so it is difficult to interpret the effects of leads and lags in this case. This result does suggest we should not focus on the releasee IV in earlier estimates, however, despite the higher F-statistic.

The effect of the IV on property crime is slightly less robust: The coefficient on the instrument in the current year is consistently negative, but the standard errors become very large and the estimates are no longer significant. As for violent crime, the signs on the lead and lag coefficients bounce between negative and positive. The only statistically significant estimate (from the specification using released convicts in the IV) suggests that the instrument from three years earlier has a negative effect on property crime in the current year. The positive coefficient on the two-year lag is also marginally-significant in several columns. Overall, no clear story about pre-existing trends (shown by the leads) emerges that would raise concerns about the property crime results.

Finally, I use the same instrumental variable strategy to test the effect of database size on the probability of arresting a suspect in newly-reported offenses. Specifically, I use equations 7 and 8, with a 0/1 indicator of whether an arrest was made as the outcome variable of interest. Standard errors are clustered at the jurisdiction (police department) level.

The expected effect is ambiguous, as the use of DNA both increases the probability of identifying repeat offenders and decreases the probability of arresting suspects without a DNA match. In addition, as crime rates decline, the composition of offenses (and offenders) is changing. Appendix Tables A-13 and A-14 present the empirical estimates of the net effect, using incident-level NIBRS data from 2000 to 2010. There are few statistically-significant estimates, though the signs of the coefficients suggest that larger databases result in a lower probability of making an arrest in new crimes. (The exception is robbery, where the coefficient is positive.) This suggestive evidence is consistent with the hypothesis that the remaining cases are harder to solve, perhaps because the easy-to-catch offenders are the first to be deterred or incapacitated by DNA profiling. The result is also consistent with the hypothesis that, as the use of DNA profiling in a particular jurisdiction increases, law enforcement is more cautious about making arrests based on traditional methods and evidence (such as eye-witness testimony). This could reduce the number of wrongful convictions.

The absence of a strong positive effect on arrest probabilities suggests that the above effects on crime rates are driven more by deterrence than by incapacitation.

5 Cost Effectiveness of DNA Databases

The value of a state DNA database depends on (1) whether the benefits of the program exceed its costs, and (2) its cost-effectiveness relative to that of other law enforcement tools such as hiring more police officers or lengthening prison sentences.

The cost of collecting and analyzing each DNA sample is currently less than $40, according to a U.S. Department of Justice estimate, and less than $20 in several states.[21] The marginal cost of analyzing new DNA samples continues to fall as technology improves, and, unlike law enforcement tools such as prisons and police officers, DNA databases exhibit tremendous returns to scale: There were large initial fixed costs in terms of crime lab equipment and computer databases, but the cost of expanding the program is relatively small.

Back-of-the-envelope calculations based on the estimated social costs of crime in McCollister, et al. (2010), are shown in Table 10 and suggest that DNA databases have resulted in dramatic savings. Based on the 95% confidence intervals on my estimates in Section 4, each profile resulted in between 0.07 and 0.68 fewer serious offenses. Multiplying these bounds by the estimated social cost of each crime gives us a social cost savings of between $1,566 and $19,945 per profile. This is a broad range, but even the most conservative estimate ($1,566) suggests DNA profiling is cost-effective. In 2010, 761,609 offender profiles were uploaded to CODIS. At $40 apiece, this cost the state and federal governments approximately $30.5 million, but saved at least $1.2 billion annually by preventing new crimes.[22] Based on my results in Section 3, much of this crime reduction was due to the deterrent effect of DNA profiling; if this holds over the longer term, it should decrease dependence on incarceration.

Owens (2009) estimates that a marginal year of incarceration results in 1.5 fewer serious offenses and costs $11,350 in Maryland. This implies that preventing a marginal crime via longer sentences costs $7,600. Estimates of the effect of police on crime rates range from 0.8

[21] See estimates at https://ncjrs.gov/pdffiles1/nij/sl000948.pdf.

[22] The Forensic Genetics Policy Initiative estimates that the fixed cost of operating and maintaining the UK's DNA database is $3 million dollars each year. If each U.S. state's fixed costs are similar, this would sum to $150 million dollars in annual fixed costs, bringing total national costs to $180.5 million. This is still far less than the $1.2 billion annual cost savings due to avoided crime.

to 1.9 fewer serious offenses per officer, per year (Levitt, 2002). Salary.com reports that the median salary for a police officer in the United States is about $50,000. This implies that preventing a marginal crime by hiring more police costs between $26,300 and $62,500, not including benefits.

In contrast, my estimates suggest that each additional DNA profile prevents at least 0.07 serious offenses, implying that the cost of preventing a marginal serious offense is, at most, $555, and falling. Based on this evidence, DNA databases are much more cost-effective, at the margin, than the most common alternatives. Note that this marginal effect does not imply that DNA profiling could substitute completely for police officers or prisons: there are undoubtedly complementarities between the three law enforcement strategies. It does suggest that a marginal dollar would be best spent on DNA profiling, or another tool that is similarly scaleable.

6 Discussion

Though DNA databases have great potential – and, anecdotally, much success – until now there has been little rigorous analysis of their effect on criminal behavior and public safety. I present evidence that DNA databases have a net deterrent effect on convicted offenders – particularly for violent offenders, but also for some property offenders – and that this individual-level effect results in a decrease in crime. The effects on crime are large, statistically significant, and economically meaningful for both violent and property offenses. This provides support for the hypothesis that it is more cost-effective to increase the probability of conviction rather than the punishment. It also provides evidence that serious offenders respond rationally to incentives not to commit crime.

The estimated benefits of DNA profiling, in terms of avoided crime, are striking. However, they do not account for privacy costs. Current law limits the use of DNA samples to creating identifying profiles, and does not allow any additional analysis of individuals' DNA. However, some worry that providing genetic material to the government gives unscrupulous analysts the opportunity to reveal health conditions and other sensitive information. In addition, legal scholars have argued that comparing offender profiles with all new crime scene evidence constitutes an unwarranted search and erodes the presumption of innocence. While such costs are difficult to quantify empirically, measuring the benefits of DNA databases helps inform public discussions of whether this tool is worth perceived costs. The estimated benefits also help us compare DNA databases with alternative high-tech tools, and invest public safety resources more effectively.

One of the biggest mysteries in the economics of crime literature is why crime has fallen across the United States over the past two decades. This study suggests that the DNA databases established and expanded during that period might be part of the answer. However, additional research is needed to tease apart DNA databases' deterrent and probative effects, and to measure the effects of adding non-felons to state databases.

References

Agan, Amanda, "Sex Offender Registries: Fear without Function?," *Journal of Law and Economics*, 2011, *54* (1), 207–239.

Ariel, Barak, William Farrar, and Alex Sutherland, "The Effect of Police Body-Worn Cameras on Use of Force and Citizens' Complaints Against the Police: A Randomized Controlled Trial," *Journal of Quantitative Criminology*, 2014.

Becker, Gary S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968, *76*, 169–217.

Bhati, Avinash, "Quantifying the Specific Deterrent Effects of DNA Databases," March 2010. Justice Policy Center, The Urban Institute.

Carr, Jillian, "The Effect of Sex Offender Registries on Recidivism: Evidence from a Natural Experiment," Available at: https://sites.google.com/site/jillianbcarr/research October 2014.

Carr, Jillian B. and Jennifer L. Doleac, "The geography, incidence, and underreporting of gun violence: new evidence using ShotSpotter data," 2015. Unpublished working paper.

Chalfin, Aaron and Justin McCrary, "Criminal Deterrence: A Review of the Literature," Working paper May 2014.

CRG Staff, "Forensic DNA, The Liberator: Interview with Peter Neufeld," *GeneWatch*, 2011, *24* (5).

DiTella, Rafael and Ernesto Schargrodsky, "Criminal Recidivism after Prison and Electronic Monitoring," *Journal of Political Economy*, 2013, *121* (1), 28–73.

Doleac, Jennifer and Nicholas Sanders, "Under the Cover of Darkness: Using Daylight Saving Time to Measure How Ambient Light Influences Criminal Behavior," *Review of Economics and Statistics*, 2015, *97* (5), 1093–1103.

Emily, Jennifer, "Dallas task force to test using DNA to solve auto thefts," *Dallas Morning News*, 2010. January 19.

Evans, William N. and Emily G. Owens, "COPS and crime," *Journal of Public Economics*, 2007, *91*, 181–201.

Hansen, Benjamin, "Punishment and Deterrence: Evidence from Drunk Driving," *American Economic Review*, 2015, *105* (4), 1581–1617.

Kilmer, Beau, Nancy Nicosia, Paul Heaton, and Greg Midgette, "Efficacy of Frequent Monitoring With Swift, Certain, and Modest Sanctions for Violations: Insights From South Dakota's 24/7 Sobriety Project," *American Journal of Public Health*, January 2013, *103* (1), e37–43.

Levitt, Steven D., "Using electoral cycles in police hiring to estimate the effects of police on crime: Reply," *American Economic Review*, 2002, *92* (4), 1244–1250.

_ , "Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not," *Journal of Economic Perspectives*, 2004, *18* (1), 163–190.

McCollister, Kathryn E., Michael T. French, and Hai Fang, "The cost of crime to society: New crime-specific estimates for policy and program evaluation," *Drug and Alcohol Dependence*, 2010, *108*, 98–109.

McGeehan, Patrick, "New York Plan for DNA Data in Most Crimes," *New York Times*, 2007. May 14.

Owens, Emily G., "More Time, Less Crime? Estimating the Incapacitative Effect of Sentence Enhancements," *Journal of Law and Economics*, 2009, *52* (3), 551–579.

_ , "Media and the Criminal Justice System," September 2010. Unpublished working paper.

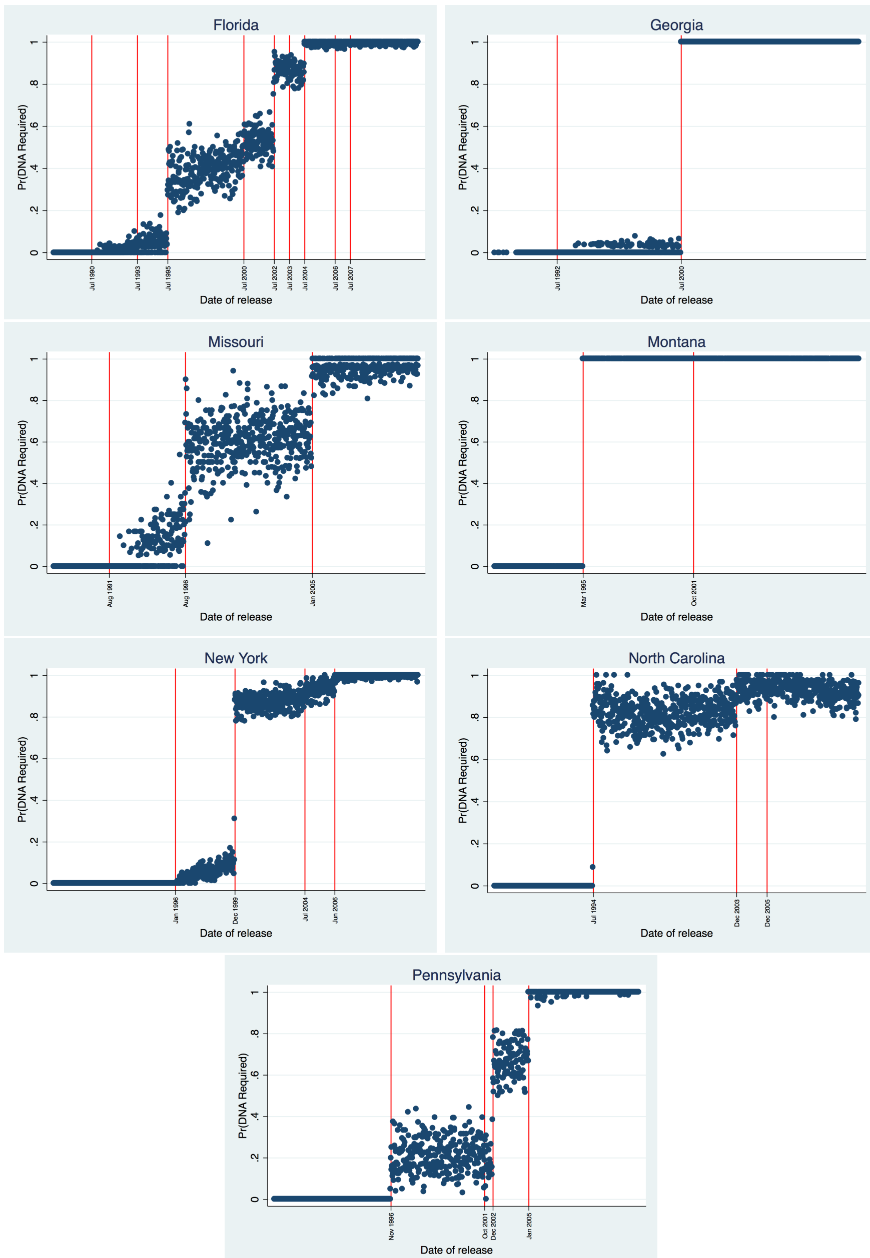
Prescott, J.J. and Jonah E. Rockoff, "Do Sex Offender Registration and Notification Laws Affect Criminal Behavior?," *Journal of Law and Economics*, 2011, *54* (1), 161–206.

Ridgeway, Greg and John MacDonald, "Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops," *Journal of the American Statistical Association*, 2009, *104* (486), 661–668.

Roman, John K, Shannon Reid, Jay Reid, Aaron Chalfin, William Adams, and Carly Knight, "The DNA Field Experiment: Cost-Effectiveness Analysis of the Use of DNA in the Investigation of High-Volume Crimes," *Urban Institute Justice Policy Center*, 2008.

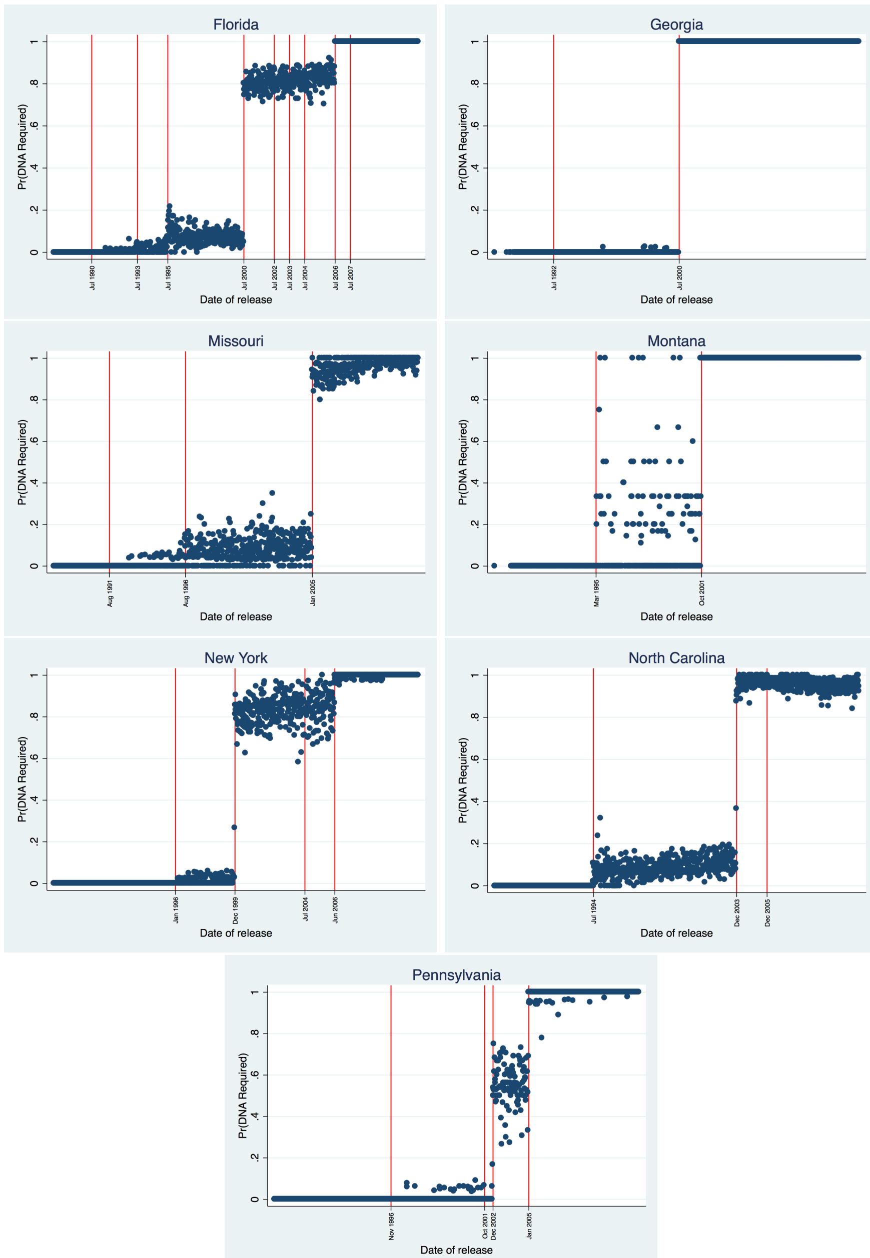
7 Figures and Tables

Figure 1: DNA requirement for serious violent offense convicts



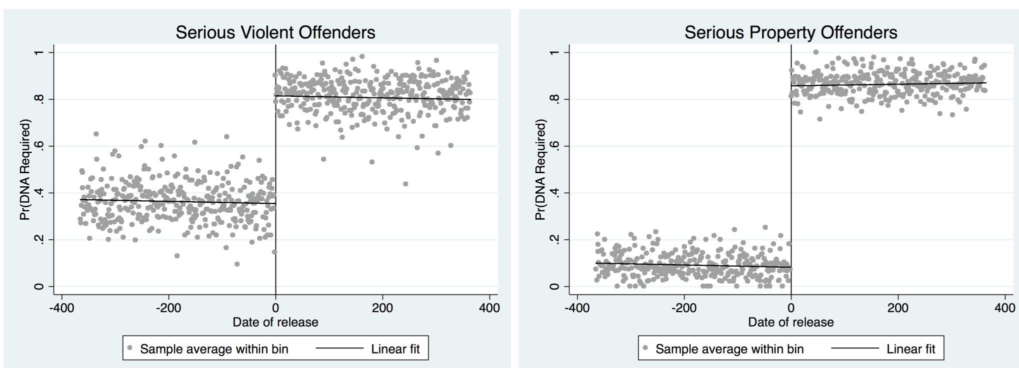
Notes: All graphs show the share of serious violent offense convicts released in a given week who were required to submit a DNA sample (based on conviction details and state law). Vertical lines show dates of database expansions. Date range: January 1, 1988, to December 31, 2011. Data source: State DOCs.

Figure 2: DNA requirement for serious property offense convicts



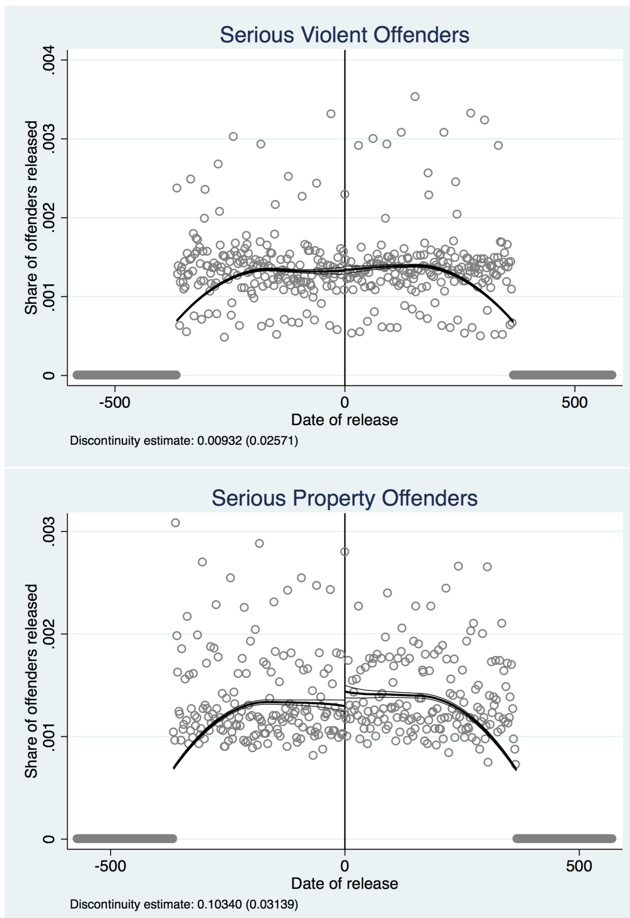
Notes: All graphs show the share of serious property offense convicts released in a given week who were required to submit a DNA sample (based on conviction details and state law). Vertical lines show dates of database expansions. Date range: January 1, 1988, to December 31, 2011. Data source: State DOCs.

Figure 3: Raw data: Probability of DNA requirement



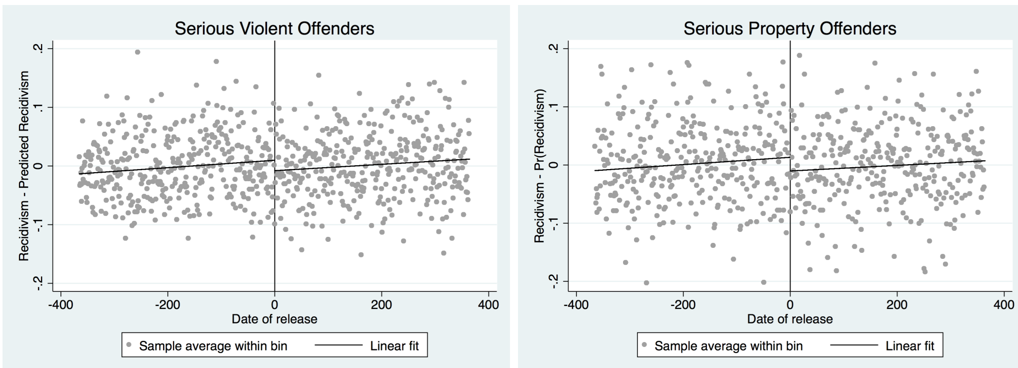
Notes: Date 0 is the date of the relevant database expansion. Bandwidth: 365 days. Data source: State DOCs and author's calculations, as described in the text.

Figure 4: McCrary test: Distribution of Released Convicts



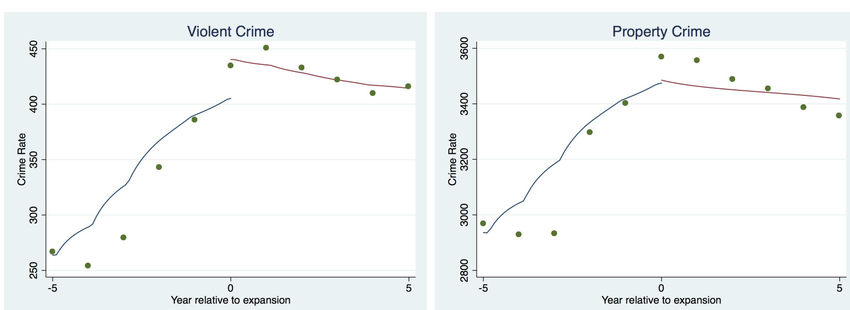
Notes: Date 0 is the date of the relevant database expansion. Bandwidth: 365 days. Data source: State DOCs.

Figure 5: Residuals: Probability of a new conviction within 5 years minus predicted probability based on observables



Notes: Bandwidth: 365 days. Date 0 is the date of the relevant database expansion. Data source: State DOCs and author's calculations, as described in the text.

Figure 6: Crime rates relative to date of "violent convict" expansion



Notes: Year 0 is the year of the relevant database expansion.

Table 1: Summary Statistics: Serious Violent Crime Convicts

Variable	Pre-Expansion (Control)					Post-Expansion (Treated)					Difference
	Mean	Std. Dev.	Min.	Max.	N	Mean	Std. Dev.	Min.	Max.	N	
DNA Required	0.363	(0.481)	0	1	24986	0.807	(0.395)	0	1	25479	0.444***
Reoffend 3 yrs	0.184	(0.387)	0	1	24986	0.192	(0.394)	0	1	25479	0.008**
Reoffend 5 yrs	0.270	(0.444)	0	1	24986	0.275	(0.447)	0	1	25479	0.005
Year of Release	2000.1	(3.357)	1993	2005	24986	2001.1	(3.378)	1994	2006	25479	0.936***
FL 1995 Expansion	0.082	(0.275)	0	1	24986	0.093	(0.290)	0	1	25479	0.011***
FL 2002 Expansion	0.180	(0.384)	0	1	24986	0.174	(0.379)	0	1	25479	-0.006*
FL 2004 Expansion	0.170	(0.376)	0	1	24986	0.175	(0.380)	0	1	25479	0.005
GA 2000 Expansion	0.066	(0.249)	0	1	24986	0.064	(0.245)	0	1	25479	-0.002
MO 1996 Expansion	0.027	(0.162)	0	1	24986	0.028	(0.164)	0	1	25479	0.001
MO 2005 Expansion	0.049	(0.215)	0	1	24986	0.050	(0.218)	0	1	25479	0.002
MT 1995 Expansion	0.004	(0.063)	0	1	24986	0.004	(0.060)	0	1	25479	-0.000
NY 1999 Expansion	0.187	(0.390)	0	1	24986	0.180	(0.384)	0	1	25479	-0.007**
NC 1994 Expansion	0.057	(0.231)	0	1	24986	0.049	(0.215)	0	1	25479	-0.008***
PA 1996 Expansion	0.035	(0.185)	0	1	24986	0.049	(0.216)	0	1	25479	0.014***
PA 2002 Expansion	0.069	(0.253)	0	1	24986	0.069	(0.253)	0	1	25479	0.000
PA 2005 Expansion	0.074	(0.262)	0	1	24986	0.066	(0.248)	0	1	25479	-0.008***
White	0.340	(0.474)	0	1	24962	0.344	(0.475)	0	1	25469	0.003
Black	0.599	(0.490)	0	1	24962	0.593	(0.491)	0	1	25469	-0.007
Age at Release	34.31	(10.50)	15.30	92.29	24519	34.41	(10.50)	15.06	92.77	25064	0.098
1st Incarceration	0.762	(0.426)	0	1	24986	0.749	(0.434)	0	1	25479	-0.013***
Incarceration Number	1.371	(0.813)	1	10	24986	1.396	(0.851)	1	10	25479	0.025***
Serious Violent Crime History	1	(0)	1	1	24986	1	(0)	1	1	25479	.
Serious Property Crime History	0.199	(0.399)	0	1	24986	0.204	(0.403)	0	1	25479	0.005
Murder History	0.082	(0.275)	0	1	24986	0.082	(0.274)	0	1	25479	-0.001
Rape History	0.079	(0.269)	0	1	24986	0.074	(0.261)	0	1	25479	-0.005**
Aggravated Assault History	0.367	(0.482)	0	1	24986	0.382	(0.486)	0	1	25479	0.015***
Robbery History	0.561	(0.496)	0	1	24986	0.555	(0.497)	0	1	25479	-0.006
Burglary History	0.143	(0.350)	0	1	24986	0.147	(0.354)	0	1	25479	0.005
Larceny History	0.061	(0.238)	0	1	24986	0.063	(0.242)	0	1	25479	0.002
Vehicle Theft History	0.055	(0.228)	0	1	24986	0.058	(0.235)	0	1	25479	0.004*
Arson History	0.005	(0.069)	0	1	24986	0.005	(0.068)	0	1	25479	-0.000
Predicted "Reoffend 3 yrs"	0.187	(0.120)	-0.293	0.520	24986	0.188	(0.121)	-0.287	0.556	25479	0.001
Predicted "Reoffend 5 yrs"	0.272	(0.157)	-0.373	0.652	24986	0.273	(0.159)	-0.368	0.680	25479	0.002

Note: Data source: Departments of Correction in Florida, Georgia, Missouri, Montana, New York, North Carolina, Pennsylvania. Sample: Convicts released before 2007 and within 365 days (before or after) DNA database expansion.

Table 2: Summary Statistics: Serious Property Crime Convicts

Variable	Pre-Expansion (Control)					Post-Expansion (Treated)					Difference
	Mean	Std. Dev.	Min.	Max.	N	Mean	Std. Dev.	Min.	Max.	N	
DNA Required	0.091	(0.288)	0	1	17187	0.864	(0.343)	0	1	17730	0.772***
Reoffend 3 yrs	0.278	(0.448)	0	1	17187	0.283	(0.450)	0	1	17730	0.005
Reoffend 5 yrs	0.381	(0.486)	0	1	17187	0.378	(0.485)	0	1	17730	-0.003
Year of Release	2000.9	(1.953)	1998	2005	17187	2001.8	(1.939)	1999	2006	17730	0.945***
FL 2000 Expansion	0.356	(0.479)	0	1	17187	0.378	(0.485)	0	1	17730	0.022***
GA 2000 Expansion	0.125	(0.330)	0	1	17187	0.122	(0.327)	0	1	17730	-0.003
MO 2005 Expansion	0.089	(0.285)	0	1	17187	0.085	(0.278)	0	1	17730	-0.004
MT 2001 Expansion	0.010	(0.102)	0	1	17187	0.012	(0.107)	0	1	17730	0.001
NY 1999 Expansion	0.123	(0.328)	0	1	17187	0.116	(0.320)	0	1	17730	-0.007**
NC 2003 Expansion	0.187	(0.390)	0	1	17187	0.182	(0.386)	0	1	17730	-0.005
PA 2002 Expansion	0.053	(0.225)	0	1	17187	0.052	(0.222)	0	1	17730	-0.001
PA 2005 Expansion	0.056	(0.231)	0	1	17187	0.054	(0.226)	0	1	17730	-0.002
White	0.504	(0.500)	0	1	17186	0.516	(0.500)	0	1	17718	0.012**
Black	0.449	(0.497)	0	1	17186	0.435	(0.496)	0	1	17718	-0.014***
Age at Release	33.13	(9.370)	15.93	78.61	16095	33.37	(9.541)	15.71	85.14	16566	0.241**
1st Incarceration	0.624	(0.484)	0	1	17187	0.621	(0.485)	0	1	17730	-0.003
Incarceration Number	1.707	(1.212)	1	20	17187	1.720	(1.232)	1	15	17730	0.013
Serious Violent Crime History	0.184	(0.388)	0	1	17187	0.179	(0.384)	0	1	17730	-0.005
Serious Property Crime History	1	(0)	1	1	17187	1	(0)	1	1	17730	.
Murder History	0.007	(0.082)	0	1	17187	0.006	(0.074)	0	1	17730	-0.001
Rape History	0.007	(0.085)	0	1	17187	0.008	(0.090)	0	1	17730	0.001
Aggravated Assault History	0.052	(0.223)	0	1	17187	0.052	(0.222)	0	1	17730	-0.000
Robbery History	0.137	(0.344)	0	1	17187	0.132	(0.339)	0	1	17730	-0.005
Burglary History	0.746	(0.435)	0	1	17187	0.754	(0.431)	0	1	17730	0.007
Larceny History	0.318	(0.466)	0	1	17187	0.306	(0.461)	0	1	17730	-0.012**
Vehicle Theft History	0.165	(0.372)	0	1	17187	0.171	(0.377)	0	1	17730	0.006
Arson History	0.024	(0.153)	0	1	17187	0.024	(0.153)	0	1	17730	-0.000
Predicted "Reoffend 3 yrs"	0.280	(0.134)	-0.156	0.663	17187	0.280	(0.133)	-0.264	0.702	17730	0.000
Predicted "Reoffend 5 yrs"	0.379	(0.153)	-0.123	0.803	17187	0.379	(0.151)	-0.261	0.812	17730	0.000

Note: Data source: Departments of Correction in Florida, Georgia, Missouri, Montana, New York, North Carolina, Pennsylvania. Sample: Convicts released before 2007 and within 365 days (before or after) DNA database expansion.

Table 3: First stage: Effect of Post-Expansion on DNA Requirement

	Serious Violent Convicts	Serious Property Convicts
Post-Expansion	0.4248***	0.7820***
	(0.0065)	(0.0063)
Release Date	-0.0003***	0.0000
	(0.0000)	(0.0000)
Black	-0.0070**	-0.0037
	(0.0032)	(0.0031)
Race missing	-0.0104	-0.0340
	(0.0589)	(0.0761)
Age	-0.0000	0.0002
	(0.0002)	(0.0002)
Age missing	0.0380**	0.0407***
	(0.0151)	(0.0089)
First incarceration	0.0257***	0.0243***
	(0.0041)	(0.0036)
Serious violent offense history		0.1469***
		(0.0040)
Serious property offense history	0.0515***	
	(0.0042)	
Murder history	0.1792***	
	(0.0074)	
Sexual Assault history	0.2848***	
	(0.0076)	
Aggravated Assault history	0.0472***	
	(0.0056)	
Robbery history	0.0361***	
	(0.0058)	
Burglary history		0.2076***
		(0.0044)
Larceny history		-0.0074*
		(0.0039)
Vehicle Theft history		-0.0855***
		(0.0046)
Arson history		-0.0245**
		(0.0101)
Constant	-0.0210***	-0.1497***
	(0.0445)	(0.0521)
Observations	50465	34917
F-statistic	4266	15415

Standard errors are shown in parentheses. * p<.10, ** p<.05, *** p<.01.
Sample: serious felony convicts released before 2007.
Bandwidth: 365 days on either side of DNA database expansion date.
State*Year FEs are included in regression but omitted from table due to limited space.

Table 4: Effect of DNA requirement on recidivism

	3-year Recidivism						**5-year Recidivism**					
	Reduced form (OLS)			Fuzzy RD (2SLS)			Reduced form (OLS)			Fuzzy RD (2SLS)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Serious Violent Convicts												
Post-Expansion	-0.0125*	-0.0124*	-0.0127*				-0.0192**	-0.0239***	-0.0241***			
	(0.0073)	(0.0075)	(0.0073)				(0.0082)	(0.0085)	(0.0083)			
Post X Release Date	-0.0000	-0.0000	0.0000				-0.0000	-0.0000	0.0000			
	(0.0000)	(0.0000)	(0.0000)				(0.0000)	(0.0000)	(0.0000)			
DNA required				-0.0294*	-0.0287*	-0.0303*				-0.0449**	-0.0554***	-0.0572***
				(0.0171)	(0.0174)	(0.0172)				(0.0191)	(0.0198)	(0.0194)
DNA X Release Date				-0.0000	-0.0000	0.0000				-0.0000	-0.0000	0.0000
				(0.0001)	(0.0001)	(0.0001)				(0.0001)	(0.0001)	(0.0001)
Observations	50465											
Pre-Expansion Mean	0.1835						0.2697					
Serious Property Convicts												
Post-Expansion	-0.0029	0.0051	0.0055				-0.0184*	-0.0093	-0.0087			
	(0.0102)	(0.0108)	(0.0106)				(0.0109)	(0.0116)	(0.0114)			
Post X Release Date	-0.0001	-0.0001	-0.0000				-0.0000	-0.0000	-0.0000			
	(0.0001)	(0.0001)	(0.0001)				(0.0001)	(0.0001)	(0.0001)			
DNA required				-0.0044	0.0058	0.0066				-0.0239*	-0.0124	-0.0115
				(0.0130)	(0.0138)	(0.0137)				(0.0140)	(0.0149)	(0.0147)
DNA X Release Date				-0.0001	-0.0001	-0.0000				-0.0000	-0.0001	-0.0000
				(0.0001)	(0.0001)	(0.0001)				(0.0001)	(0.0001)	(0.0001)
Observations	34917											
Pre-Expansion Mean	0.2775						0.3809					
Controls:												
State X Year FEs	X	X	X	X	X	X	X	X	X	X	X	X
Month FEs		X	X		X	X		X	X		X	X
Demographic			X			X			X			X
Criminal History			X			X			X			X

Note: Standard errors are shown in parentheses. * p<.10, ** p<.05, *** p<.01. Coefficients show the marginal effect of DNA requirement on the probability of a subsequent conviction. Data source: Departments of Correction in Florida, Georgia, Missouri, Montana, New York, North Carolina, Pennsylvania. Sample: Convicts released before 2007 and within 365 days (before or after) DNA database expansion. Basic specification includes: state FEs, year FEs, running variable (recentered released date). Demographic controls include: black, race missing, age, age missing. Criminal history controls include: first incarceration, ever convicted of: (for serious violent convicts) serious property crime, murder, sexual assault, aggravated assault, robbery; (for serious property convicts) serious violent crime, burglary, larceny, vehicle theft, arson.

Table 5: Effect of DNA requirement on recidivism: Violent offenders

	3-year Recidivism Fuzzy RD (2SLS)		**5-year Recidivism** Fuzzy RD (2SLS)	
	(1)	(2)	(3)	(4)
Murder History				
DNA collection	0.0009	0.0083	-0.0176	0.0219
	(0.0653)	(0.0618)	(0.0742)	(0.0711)
DNA X Release Date	-0.0003	-0.0002	-0.0003	-0.0002
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Observations	4777			
Pre-Expansion Mean	0.1140		0.1626	
Rape History				
DNA collection	-0.0604	-0.0834	-0.0832	-0.1173
	(0.0816)	(0.0773)	(0.0918)	(0.0873)
DNA X Release Date	-0.0006	-0.0006	-0.0004	-0.0003
	(0.0004)	(0.0004)	(0.0005)	(0.0004)
Observations	4521			
Pre-Expansion Mean	0.1791		0.2608	
Aggravated Assault History				
DNA collection	-0.0296	-0.0472	-0.0159	-0.0396
	(0.0360)	(0.0381)	(0.0394)	(0.0426)
DNA X Release Date	0.0001	0.0001	0.0003	0.0003
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Observations	24774			
Pre-Expansion Mean	0.2258		0.3258	
Robbery History				
DNA collection	-0.0338	-0.0362	-0.0721**	-0.1024***
	(0.0288)	(0.0295)	(0.0309)	(0.0327)
DNA X Release Date	0.0001	0.0000	-0.0000	-0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Observations	39467			
Pre-Expansion Mean	0.2478		0.3535	
Controls:				
State X Year FEs	X	X	X	X
Month FEs		X		X
Demographic		X		X
Criminal History		X		X

Note: Standard errors are shown in parentheses. * p<.10, ** p<.05, *** p<.01. Coefficients show marginal effect of DNA requirement on the probability of a subsequent conviction. Data source: Departments of Correction in Florida, Georgia, Missouri, Montana, New York, North Carolina, Pennsylvania. Sample: Convicts released before 2007 and within 365 days (before or after) DNA database expansion. Basic specification includes: state FEs, year FEs, running variable (recentered released date). Demographic controls include: black, race missing, age, age missing. Criminal history controls include: first incarceration, ever convicted of: serious property crime, murder, sexual assault, aggravated assault, robbery.

Table 6: Effect of DNA requirement on recidivism: Property offenders

	3-year Recidivism Fuzzy RD (2SLS)		5-year Recidivism Fuzzy RD (2SLS)	
	(1)	(2)	(3)	(4)
Burglary History				
DNA collection	-0.0126	-0.0022	-0.0308*	-0.0188
	(0.0149)	(0.0159)	(0.0159)	(0.0169)
DNA X Release Date	-0.0000	0.0000	-0.0000	0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Observations	37191			
Pre-Expansion Mean	0.3119		0.4210	
Larceny History				
DNA collection	-0.0443	-0.0517*	-0.0900***	-0.0871***
	(0.0283)	(0.0307)	(0.0299)	(0.0324)
DNA X Release Date	-0.0001	-0.0001	-0.0001	-0.0000
	(0.0002)	(0.0001)	(0.0002)	(0.0002)
Observations	16818			
Pre-Expansion Mean	0.3328		0.4466	
Vehicle Theft History				
DNA collection	-0.0316	-0.0541	0.0103	-0.0438
	(0.0581)	(0.0767)	(0.0602)	(0.0804)
DNA X Release Date	-0.0007*	-0.0005	-0.0006*	-0.0005
	(0.0003)	(0.0003)	(0.0004)	(0.0003)
Observations	8179			
Pre-Expansion Mean	0.3546		0.4783	
Arson History				
DNA collection	0.0044	-0.0154	0.1016	0.1060
	(0.1485)	(0.1311)	(0.1738)	(0.1536)
DNA X Release Date	0.0001	0.0001	-0.0004	-0.0004
	(0.0006)	(0.0006)	(0.0007)	(0.0007)
Observations	1049			
Pre-Expansion Mean	0.1721		0.2428	
Controls:				
State X Year FEs	X	X	X	X
Month FEs		X		X
Demographic		X		X
Criminal History		X		X

Note: Standard errors are shown in parentheses. * p<.10, ** p<.05, *** p<.01. Coefficients show marginal effect of DNA requirement on the probability of a subsequent conviction. Data source: Departments of Correction in Florida, Georgia, Missouri, Montana, New York, North Carolina, Pennsylvania. Sample: Convicts released before 2007 and within 365 days (before or after) DNA database expansion. Basic specification includes: state FEs, year FEs, running variable (recentered released date). Demographic controls include: black, race missing, age, age missing. Criminal history controls include: first incarceration, ever convicted of: serious violent crime, burglary, larceny, vehicle theft, arson.

Table 7: Summary Statistics: Crime Rate Analysis

Variable	Mean	Std. Dev.	Min.	Max.	N
Crime Rates					
Violent Crime rate	399.68	165.28	77.780	822.00	344
Property Crime rate	3179.6	787.19	1619.6	5632.4	344
Murder rate	4.6156	2.3327	0.4677	13.229	344
Rape rate	32.868	10.746	11.158	92.476	344
Assault rate	258.67	122.18	42.579	626.46	344
Robbery rate	103.63	56.562	6.1479	256.73	344
Burglary rate	668.34	239.38	292.30	1216.1	344
Larceny rate	2210.0	508.12	1188.9	3562.9	344
Vehicle Theft rate	301.26	150.55	73.530	1021.3	344
Years of Database Expansions					
Sex Convicts	1994.3	2.7936	1988	1999	344
Violent Convicts	1996.8	3.5835	1989	2005	344
Burglary Convicts	2000.8	3.4898	1990	2009	344
All Felony Convicts	2003.0	3.9314	1990	2010	333
Sex Inmates	2000.8	1.8719	2000	2008	274
Violent Inmates	2001.0	1.8936	2000	2008	274
Burglary Inmates	2002.1	2.1998	2000	2008	243
All Felony Inmates	2003.7	2.7975	2000	2010	227
Database Size					
DNA Profile rate, 2000	259.16	388.23	3.5659	1904.7	22
DNA Profile rate, 2001	329.76	471.19	21.945	2349.6	22
DNA Profile rate, 2002	507.37	537.79	51.824	2563.9	23
DNA Profile rate, 2003	548.60	472.44	0.9361	2733.4	47
DNA Profile rate, 2004	688.56	640.64	20.421	3050.1	25
DNA Profile rate, 2005	873.80	723.03	37.559	3199.6	27
DNA Profile rate, 2006	1374.9	959.36	87.583	3359.2	21
DNA Profile rate, 2007	1684.3	949.6	120.55	3493.8	19
DNA Profile rate, 2008	1777.2	853.90	213.03	3778.3	46
DNA Profile rate, 2009	2232.5	923.15	304.18	3920.5	46
DNA Profile rate, 2010	2442.1	1047.5	356.97	4137.7	46
DNA Profile rate, All Years	1304.9	1101.5	0.9361	4137.0	344
Instrumental Variables					
IV – reported crime	18369	16592	44.280	71551	344
IV – arrests	4537.8	3550.0	23.856	16054	344
IV – arrests & under 40	3613.9	2805.9	19.085	12701	344
IV – convicts	2774.9	2240.3	15.770	10032	344
IV – convicts & under 40	2187.9	1739.7	12.707	7817.5	344
IV – released convicts	3851.5	2547.7	104.50	12042	344
IV – released convicts & under 40	3039.1	1981.0	83.591	9403.3	344
Police Officer rate	227.21	50.690	155.08	405.75	344

Note: Crime, police and DNA profile statistics are per 100,000 residents. Data sources: (1) Crime rates: FBI UCR; (2) years of legislative expansion: legislative histories; (3) police officers: FBI LEOKA: (4) DNA profiles: a variety of sources, as described in the Appendix; (5) simulated instrument: author's calculation, as described in text.

Table 8: First Stage: Effect of Instrument on Profiles

	DNA Profiles in Database
Predicted qualifying offenders	0.4163***
	(0.0777)
Observations	344
F-statistic	28.679

Note: Coefficient shows the effect of the simulated IV (predicted qualifying offenders, based on number of convicts) on database size. Standard errors in parentheses.
* p<.10, ** p<.05, *** p<.01. Specification includes year fixed effects, state fixed effects, and police officers per capita. Standard errors are clustered by state. Simulated instrument and profiles are per 100,000 state residents.

Table 9: Effect of DNA Database Size on Crime Rates

	Effect of a marginal DNA profile per 100,000							
	OLS	Simulated IV						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Violent Crime Rate	-0.0139	-0.0466**	-0.0530***	-0.0535***	-0.0511***	-0.0520***	-0.0480***	-0.0490***
	(0.0090)	(0.0186)	(0.0189)	(0.0189)	(0.0188)	(0.0188)	(0.0166)	(0.0167)
2000 mean	*427.06*							
Property Crime Rate	-0.0667	-0.3281**	-0.3214***	-0.3215***	-0.3231**	-0.3230***	-0.3192***	-0.3197***
	(0.0937)	(0.1407)	(0.1212)	(0.1198)	(0.1264)	(0.1238)	(0.1150)	(0.1129)
2000 mean	*3553.2*							
Murder Rate	-0.0000	-0.0004***	-0.0004**	-0.0004**	-0.0004**	-0.0004**	-0.0003**	-0.0003**
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
2000 mean	*4.7760*							
Rape Rate	-0.0010	-0.0044**	-0.0045***	-0.0046***	-0.0045***	-0.0045***	-0.0042***	-0.0043***
	(0.0009)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0015)	(0.0015)
2000 mean	*33.981*							
Assault Rate	-0.0106	-0.0307**	-0.0364**	-0.0368**	-0.0347**	-0.0354**	-0.0329**	-0.0337**
	(0.0074)	(0.0152)	(0.0155)	(0.0155)	(0.0154)	(0.0154)	(0.0138)	(0.0139)
2000 mean	*280.35*							
Robbery Rate	-0.0023	-0.0110***	-0.0117***	-0.0117***	-0.0115***	-0.0116***	-0.0106***	-0.0107***
	(0.0017)	(0.0032)	(0.0031)	(0.0031)	(0.0031)	(0.0031)	(0.0025)	(0.0025)
2000 mean	*107.96*							
Burglary Rate	0.0024	-0.0208	-0.0209	-0.0208	-0.0210	-0.0208	-0.0195	-0.0194
	(0.0168)	(0.0236)	(0.0215)	(0.0213)	(0.0220)	(0.0217)	(0.0197)	(0.0195)
2000 mean	*703.22*							
Larceny Rate	-0.0388	-0.2151**	-0.2060**	-0.2046**	-0.2087**	-0.2066**	-0.2042**	-0.2026**
	(0.0641)	(0.0990)	(0.0850)	(0.0841)	(0.0887)	(0.0868)	(0.0812)	(0.0798)
2000 mean	*2496.1*							
Vehicle Theft Rate	-0.0302	-0.0923***	-0.0945***	-0.0960***	-0.0934***	-0.0956***	-0.0955***	-0.0977***
	(0.0183)	(0.0250)	(0.0224)	(0.0223)	(0.0230)	(0.0227)	(0.0210)	(0.0210)
2000 mean	*353.88*							
F-statistic		22.325	30.565	31.452	28.679	30.420	50.058	50.943
Observations	344							
Instrument:								
IV = reported crimes		X						
IV = arrests			X	X				
IV = convicts					X	X		
IV = released convicts							X	X
IV = Under age 40				X		X		X

Note: Standardized standard errors in parentheses. * p<.10, ** p<.05, *** p<.01. Each coefficient indicates the change in the crime rate (per 100,000 residents) resulting from an increase in the number of DNA profiles (per 100,000 residents) in the state database. Instrumental variables are the simulated stock and flow of qualifying offenders. Crime rate and police officer data source: FBI UCR. Sentenced prisoner data source: BJS, Prisoner series.

Table 10: Effect of DNA Databases on the Annual Social Cost of Crime

	Reported Offenses per DNA Profile 95% Confidence Interval		Social Cost per	Social Cost per DNA Profile	
	Lower Bound	Upper Bound	Offense	Lower Bound	Upper Bound
Murder	-0.0008	-0.0000	$8,982,907	-$7,114	-$72
Rape	-0.0078	-0.0012	$240,776	-$1,886	-$281
Assault	-0.0649	-0.0045	$107,020	-$6,944	-$483
Robbery	-0.0176	-0.0054	$42,310	-$744	-$229
Burglary	-0.0641	0.0221	$6,462	-$414	$143
Larceny	-0.3826	-0.0348	$3,532	-$1351	-$123
Vehicle Theft	-0.1385	-0.0483	$10,772	-$1492	-$521
Any Serious Offense	-0.6763	-0.0721		-$19,945	-$1,566

Estimates are based on the coefficients from Column 5 of Table 9 and the estimated social costs of crimes presented in McCollister, et al. (2010).