Online Appendix to *Better Schools, Less Crime?*

David J. Deming

July 2011

A.1 Sample and Data Sources

The analysis sample consists of 44,028 students in grades 6 through 11 who were enrolled in any CMS school in the previous year. These students listed as first choices 28 different middle schools and 17 different high schools. 26,474 students listed first a school to which they were guaranteed admission. Of the remaining 17,554 students, 5,033 were in lotteries where no students were offered admission, and 8,310 were in lotteries where all students were accepted. This left 4,211 students with admission to a first choice school that was subject to randomization (1,891 in high school and 2,320 in middle school). Nearly all schools had some applicants that were randomized (24 of the 28 middle schools, and 16 of the 17 high schools). Together with different priority groupings for grades and free lunch-eligible applicants, there were 72 lotteries in the middle school sample and 34 lotteries in the high school sample. About 46% of high school lottery applicants and 38% of middle school lottery applicants were admitted to their first choice school, although this varied tremendously by lottery.

The lottery file comes from Hastings et al (2008) and includes students' individual choices, priority groupings, and lottery numbers. Within each priority group, lottery numbers were randomly assigned to students and slots were filled in ascending order by lottery number. I verified that the lottery numbers were accurate by plotting the probability of enrollment against within-priority-group lottery numbers and looking for evidence of a sharp break in enrollment at the minimum number cutoff. These graphs are available on request.

A.1.1 CMS Administrative Data

CMS maintains yearly student records that are linked longitudinally with a unique student identification number. The North Carolina Department of Public Instruction (NCDPI) requires CMS to report end-of-year (EOY) files for each school and grade with student enrollment, demographics, behavior measures and yearly test scores in a standard format. In addition to basic demographic information, these files include standardized math and reading End-of-Grade (EOG) tests for grades 3 through 8, End-of-Course (EOC) exams scores for specific subjects (such as Algebra I, Chemistry, and English I) taken mostly in high school, excused and unexcused absences, total days out-of-school suspended, special education classifications (with information about the nature and severity of the disability) and limited English proficiency status.

In addition to these EOY files, I have obtained more detailed information under a data use agreement with CMS and the Harvard Center for Education Policy Research (CEPR). The data are stored on secure computers with no internet connectivity in a room at CEPR. Access is restricted to identified researchers by means of a keycard system. The data include student's name, date of birth, and exact address. They also include yearly course enrollment information and grade received, which I can use to construct measures of grade point average and accumulated credits. I use address information to group students into census tract-by-school zone “neighborhoods”, and I control for these neighborhood fixed effects in the crime prediction regression in Section III.A. Following Hastings et al (2008), I also use address information to calculate straight-line distance from each student's home to each school, which I use in the revealed preference calculation in Table I.

The CMS administrative data also contains dates of school enrollment and withdrawal. Each spell of enrollment has an associated withdrawal code. Withdrawal codes include high school graduation, transfer within CMS, transfer to private or charter schools, transfer to another public school in-state, out-of-state transfer, dropout, and no show, as well as other categories such as assignment to alternative schools, expulsion and death. CMS also provided a teacher information file, which includes courses taught, years of experience and information about the colleges attended and degrees obtained. I match each teacher's undergraduate institution to the Barron's Profile of American Colleges 2009, which groups schools into categories such as “competitive”, “very competitive, and “most competitive”, and use these classifications in the measure of teacher quality in Table I.

A.2 Crime Data Collection and Match Process

Arrest data at the county level come from the Mecklenburg County Sheriff. The data include all arrests made in Mecklenburg county, including by arresting agencies with other jurisdictions (ex. Immigration and Naturalization Services, the US Marshals and other federal agencies, as well as city police from Charlotte and surrounding smaller cities). The data include all arrests made beginning on January 1st, 2006 through June 15th, 2009, with the exception of the approximately 3% of arrests that were expunged or missing. The data are collected at the arrest level, and include information on the classification (felony, misdemeanor, traffic), processing (bond amount, warrant, etc.) and exact description all associated charges at the time of arrest. Each arrest is assigned a unique 7 digit number in the order that it is processed, and first time arrestees are assigned a unique 6 digit identification number (established by fingerprinting) that links them across multiple arrests, if any. I have information on each arrestee's name and date of birth, which I use to match to the CMS administrative data, as well as home address at the time of arrest. MCS incarceration data cover the same period of time as the arrest data and are kept in a similar format. The unique 6 digit identification number links individuals to all spells of incarceration in MCS jails, and the associated charges. The data include name and date of birth and the first and last day of each incarceration spell.

The original source for the 2006-2009 Mecklenburg county arrest and incarceration data is http://www.charmeck.org/Departments/MCSO/Inmate+Information/InmateLookup.htm. As the website states, “North Carolina Law makes this information public. The Mecklenburg County Sheriff's Office provides it via the internet for your convenience.” The arrest data can be found at http://arrestinquiryweb.co.mecklenburg.nc.us/ and the incarceration data at http://mcsowebsvr.co.mecklenburg.nc.us/inmatesearch/inmate_search.asp. Both websites allow users to access information that is up to 3 years old, counting from the day the website is accessed (since I started collecting the data on January 1st, 2009, my data begin on January 1st, 2006). I collected the data by writing a script (also known as a macro) in an automation language called AutoIt. This program, which is similar to the more commonly used Perl, allows me to automate keystrokes, mouse clicks and other basic computer functions. MCS assigns arrest numbers consecutively in the order they are processed, so I wrote a script that entered arrest numbers in order into the website and copied all the relevant information into a text file. The websites both include name and date of birth, so I was able to connect arrests to individuals, and then individual arrestees (in some cases) to student records in CMS. Because of the format of the website, I was unable to fully automate collection of the incarceration data. Therefore, I collected incarceration data for African-American members of the lottery sample only.

I also obtain data from the North Carolina Department of Corrections (NCDOC). These data include spells of incarceration and associated charges and convictions for individuals who serve time in state prison. Members of the lottery sample can thus be linked to crimes committed outside of Mecklenburg county, but only if they spend time in state prison for those crimes. The NCDOC data include spells of incarceration prior to 2006, but only for individuals who are incarcerated or under the supervision of the justice system (i.e. on probation) as of 2009. Data from 2006 to the present do not have this limitation. Therefore, I also limit analysis of the NCDOC incarceration data to 2006 and later, for consistency. Like the MCS incarceration data, I was unable to fully automate collection of the NCDOC data, so I restrict to African-American members of the lottery sample only. Finally, I matched the crime data to CMS administrative data using first name, last name, and exact date of birth. To account for inconsistencies across data sources (i.e. hyphenated names, apostrophes, “Dave” vs. “David” etc.) I employed a partial matching algorithm. I used a STATA program written by Eric Taylor at CEPR called “lndmerge” that calculates the Levenshtein distance between two variables using optimal matching of sequences. The procedure is as follows: first the matching variables in each data source (i.e. name and date of birth) are combined into a unique string. Then all the observations in both datasets are combined into a matrix, and each combination is assigned a score (or distance) based on how many changes would need to be made to obtain an exact match. Longer strings are less likely to be exact matched, and so are penalized proportionately less for a change (i.e. David-Devid would count as a worse match than DavidDeming-DevidDeming). Using this method, about 87% of the matches were exact. I adopted various rules for accepting partial matches (a minimum score, minimum score plus exact match on first letter of last name, or on year of birth etc.) None of these made any difference in the main results, nor did restricting the analysis to exact matches only.

I conducted a number of tests to assess the quality of the match. First, since each arrest is given a unique identification number that is assigned consecutively in the order it was processed, I can calculate the fraction of arrest numbers that are missing from the data. Counting from the first day that the data were collected, this fraction is only 3.2%, and there are no large gaps. This suggests that nearly every arrest processed by MCS is present in the data. Online Appendix Figure I plots the age profile of arrests in Mecklenburg County by type of offense. The Federal Bureau of Investigation (FBI) collects data on eight different “index” crimes for the Uniform Crime Reporting (UCR) Program, which covers law enforcement agencies across the country. Index property crimes are burglary, motor vehicle theft and felony larceny. Index violent crimes include murder/manslaughter, rape, robbery and aggravated assault. The last category I include is felony drug offenses, which (based on weight) range from "possession with intent to distribute" all the way up to "trafficking." Index property and violent crimes peak at ages 17 and 18 respectively, which is consistent with other cohort studies of crime and delinquency (Wolfgang et al 1987, Farrington et al 1986, Sampson and Laub 2003). Interestingly, drug felony arrests do not peak until the early to mid-twenties, and decline much more slowly with age than other categories of crime.

In the top panel of Online Appendix Table I, I examine arrest rates of CMS attendees overall and by demographic group. I use six school cohorts of data, corresponding to students in grades 6 through 11 in 2002 and age 17 to 23 in 2009. The first and second rows show the fraction of CMS attendees who have a criminal record, and who have at least one felony arrest respectively, by race and gender. Not surprisingly, arrest rates vary dramatically, from about 34% for African-American males to about 3% for White or Asian females. Rows three through five show arrest rates by type of crime. African-American males are about six times more likely than white males to have at least one felony arrest, and about thirteen times more likely to be arrest for an index violent crime.

In the bottom panel of Online Appendix Table I, I examine the percentage of arrests that are successfully matched to a CMS student by birth year and demographic group. Unmatched arrests could be students who were enrolled in private school, youth who travel to Mecklenburg County from elsewhere to commit crimes, or poor data quality. Match rates are highest for African-Americans (who are more likely than whites to attend public school) and for more recent birth years. Since the CMS data only go back to the 1996-1997 school year, any student who left the district before that would not be matched. Since most criminals are high school dropouts, this is likely to result in fewer matches for the earliest birth cohorts. However, the weighted average match rate by birth year for the lottery sample exceeds 85% overall and 90% for African-American males. This high match rate is strong evidence of the quality of the data. It also highlights the important role that public school policies might play in city crime rates.

A.3 Selection into the Lottery Sample

Online Appendix Table II presents the average characteristics of lottery applicants compared to all CMS students. Column 1 shows control means and Column 2 shows coefficients from regressions of observable characteristics of students on an indicator for whether the student listed a non-guaranteed school as their first choice. Unlike many other instances of school choice, applicants to non-guaranteed schools are more disadvantaged than students who choose their neighborhood school. They are nearly twice as likely to be nonwhite and free or reduced price lunch eligible. Applicants to non-guaranteed schools also score about 0.4 standard deviations lower on both math and reading exams, and have been suspended and absent more days in the previous school year. Column 3 includes neighborhood school fixed effects, to assess the nature of within-school selection. Column 4 presents control means and Column 5 presents estimates where the sample is restricted to neighborhood schools where 60% or more of the assigned students are African-American or Latino.

Although applicants to non-guaranteed schools are more disadvantaged across schools, they are relatively similar on observables within the schools from which most of the lottery sample comes. Column 5 shows that, even with predominately minority schools, non-guaranteed applicants have test scores that are very similar to students who chose the neighborhood school. Furthermore, even within these high minority schools, applicants to non-guaranteed schools are absent and suspended more often. Column 6 looks only at students who were in non-degenerate lotteries (where the probability of admission was neither zero nor one). We see that applicants in the lottery sample have slightly higher test scores (about 0.1 standard deviations). However, this is largely because of the “priority boost” given to economically disadvantaged applicants, many of whom were automatically admitted and thus not subject to randomization. Overall, the lottery sample is more disadvantaged than the average CMS student, but quite representative on observables of the students who attend high minority schools.

A.4 Arrest Prediction

I estimate the probability that a student will have at least one arrest as a function of yearly test scores in math and reading, absences and out-of-school suspensions, special education classifications, and neighborhood school zone by census tract fixed effects using each student’s exact address in the year prior to open enrollment. For the high school sample I use data from grades 6 through 8, and grades 3 through 5 for the middle school sample. I allow for second order polynomials in all of the continuous measures. The coefficients from the regression are listed in Online Appendix Table III. In Columns 3 and 4, I reestimate the model with males only.

These coefficients, which are the ones actually used in the crime prediction for the main results, differ very little from the prediction for the overall sample. Online Appendix Figure II plots the density of predicted criminality for all CMS students in grades 6 to 11, then for African-American males overall and from the seven lowest-performing schools (defined by average test scores) in the district. The distribution shifts rightward noticeably for these “high risk” subgroups.

A.5 Social Cost of Crime Calculations

The social cost of crime estimates Miller et al (1996) include tangible costs such as lost productivity, medical and mental health care and other social services, and property damage. They also include estimates of intangible costs such as quality of life (based in part on the amount individuals are willing to pay to reduce the risk of death, and the compensatory component of jury damage awards - see Miller et al (1996) for details). Intangible costs make up most of the estimated cost of violent crimes, and are inherently difficult to monetize. Notably, the study does not include criminal justice system costs such as policing, crime and arrest processing, or incarceration. It also does not include the costs undertaken by individuals to avoid crime. Here I list the costs for the index property and index violent crimes, plus a few other notable crimes that drive the main estimates in the paper (all estimates are converted to 2009 dollars).

1. Murder - $4.38 million
2. Rape - $129,630
3. Aggravated Assault - $35,760
4. Domestic Assault - $16,390
5. Simple Assault - $2,980
6. Robbery - $11,920
7. Motor Vehicle Theft - $5,513
8. Burglary - $2,086
9. Larceny - $551

Miller et al (1996) do not monetize all crimes, and notably they exclude drug crimes from the estimation. One alternative is to impute a cost of zero for all drug crimes. This leaves the estimates for the middle school sample unchanged, but reduces the social cost estimates for the high school sample by approximately 25%. In the main estimates in the paper, I impute a cost of drug felonies that is equivalent to felonies of the same standing under the North Carolina Structured Sentencing Act. This varies by crime and the "schedule" of the controlled substance (for example, cocaine is schedule 2 and punished more severely than marijuana, which is schedule 6). The approximate classifications are below (for marijuana, crimes are roughly one step down in severity, so trafficking in marijuana = sell/deliver cocaine, roughly):

1. Drug Trafficking = Robbery = $11,920
2. Sell/Deliver = Motor Vehicle Theft = $5,513
3. Possession with Intent to Distribute = Burglary = $2,086
4. Simple Possession (Felony) = Larceny = $551

Online Appendix Tables and Figures

Appendix Table I: Arrest Rates and Match Between School District and Arrest Data

Panel A: Arrest Rates by Race/Gender and Crime Type

	African-American		Hispanic		White/Asian		
	Male	Female	Male	Female	Male	Female	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ever Arrested							
Any Arrest	0.34	0.13	0.16	0.04	0.10	0.03	
Any Felony	0.20	0.03	0.08	0.01	0.03	0.01	
Index Property	0.09	0.01	0.04	0.00	0.01	0.00	
Index Violent	0.07	0.00	0.02	0.00	0.01	0.00	
Drug Felony	0.08	0.01	0.01	0.00	0.01	0.00	
Sample Size	8,834	8,493	519	504	9,095	8,748	

Panel B: Percent of Arrests Matched to a CMS Attendee

	African-American		Hispanic		White/Asian		All
Year of Birth	Male	Female	Male	Female	Male	Female	Felonies
1980	0.26	0.20	0.01	0.00	0.11	0.04	0.19
1981	0.59	0.39	0.02	0.03	0.27	0.22	0.44
1982	0.65	0.56	0.03	0.08	0.34	0.25	0.53
1983	0.73	0.73	0.03	0.09	0.43	0.33	0.62
1984	0.72	0.66	0.04	0.09	0.48	0.42	0.64
1985	0.79	0.76	0.08	0.04	0.49	0.42	0.70
1986	0.83	0.74	0.12	0.24	0.53	0.43	0.75
1987	0.85	0.78	0.13	0.24	0.59	0.53	0.80
1988	0.90	0.86	0.23	0.31	0.72	0.67	0.85
1989	0.93	0.88	0.40	0.76	0.73	0.71	0.89
1990	0.93	0.91	0.57	0.75	0.82	0.68	0.90
1991	0.94	0.92	0.79	0.88	0.80	0.81	0.91
1992	0.95	0.94	0.74	0.83	0.81	0.80	0.91
1993	0.97	0.82	0.75	1.00	0.80	0.57	0.95
All Years	0.77	0.72	0.13	0.22	0.49	0.42	0.69
Sample Size	32,598	7,459	10,392	715	12,161	4,085	19,184

Notes: The sample in panel A consists of CMS attendees in grades K-5 in 1997 (ages 17-23 in 2009) that are still in CMS in grade 8 or higher. Index property crimes are felony larceny, burglary and motor vehicle theft. Index violent crimes are murder/manslaughter, aggravated assault, robbery and kidnapping. In Panel B the denominator is all arrests in Mecklenburg County.

Appendix Table II: Selection into the Lottery Sample

	Outcome - Chose Non-Guaranteed School					
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.51	-0.01	-0.01	0.50	0.00	-0.00
		[0.01]	[0.01]		[0.01]	[0.01]
African-American or Latino	0.40	0.27***	0.13***	0.73	0.03	0.04
		[0.01]	[0.03]		[0.03]	[0.03]
Free / Reduced Lunch	0.40	0.26***	0.12***	0.71	0.04	0.01
		[0.01]	[0.02]		[0.02]	[0.03]
Math (standardized)	0.15	-0.41***	-0.16***	-0.36	-0.03	0.13***
		[0.01]	[0.01]		[0.04]	[0.04]
Reading (standardized)	0.15	-0.41***	-0.16***	-0.37	-0.04	0.11**
		[0.01]	[0.04]		[0.05]	[0.05]
Days Suspended	0.63	0.55***	0.36***	0.99	0.33***	0.04
		[0.04]	[0.08]		[0.13]	[0.08]
Days Absent	7.32	1.48***	1.02***	7.94	0.97***	0.37
		[0.09]	[0.19]		[0.31]	[0.30]
Home School FE			X	X	X	X
>60% Nonwhite Only				X	X	X
Non-Degenerate Lotteries Only						X
Sample Size	44,028			18,353		

Notes : The sample is all CMS students in rising grades 6-11 in the fall of 2002 w ho w ere enrolled in any CMS school in the previous year. The first column presents the control mean and the second column presents coefficients from a regression of the variable in each row on an indicator for w hether the student listed a non-guaranteed school as their first choice. The third column adds neighborhood school fixed effects. Columns 4 show s the control mean and Column 5 show s estimates w hen the sample is restricted to schools w here the assigned student population is 60% or more nonw hite. In Column 6 the independent variable of interest is an indicator for w hether the student w as in the lottery sample (i.e. they w ere in a priority group w here the probability of admission w as neither zero nor one.) Free or reduced price lunch is an indicator of socioeconomic status. Math and Reading are standardized scores administered in the years that students w ere in 5th grade (for middle school) and 8th grade (for high school). Standard errors are clustered at the neighborhood school level. * - sig. at 10% level. ** - sig. at 5% level. *** - sig. at 1% level.

Appendix Table III: Arrest Prediction

Dependent Variable: Ever Arrested (Logit Coefficients)

	All		Males Only	
	High (6-8 Xs)	Middle (3-5 Xs)	High (6-8 Xs)	Middle (3-5 Xs)
Demographics				
Male	**1.16 (0.05)**	**0.93 (0.05)**		
Black	**0.47 (0.07)**	**0.41 (0.07)**	**0.50 (0.08)**	**0.41 (0.08)**
Latino	**-0.70 (0.16)**	**-0.29 (0.11)**	**-0.60 (0.18)**	**-0.24 (0.13)**
FRPL	**0.32 (0.07)**	**0.47 (0.07)**	**0.31 (0.08)**	**0.37 (0.08)**
Math Scores				
6th / 3rd	-0.05 (0.07)	0.03 (0.06)	-0.03 (0.08)	0.03 (0.08)
squared	0.02 (0.03)	-0.02 (0.03)	0.02 (0.04)	-0.03 (0.03)
7th / 4th	-0.05 (0.07)	-0.01 (0.06)	-0.11 (0.09)	-0.05 (0.08)
squared	-0.00 (0.03)	0.01 (0.03)	-0.04 (0.04)	0.00 (0.03)
8th / 5th	-0.10 (0.07)	**-0.19 (0.06)**	-0.05 (0.08)	**-0.23 (0.07)**
squared	-0.05 (0.03)	-0.00 (0.02)	-0.04 (0.04)	0.03 (0.03)
Reading Scores				
6th / 3rd	**-0.14 (0.07)**	-0.09 (0.06)	-0.13 (0.08)	**-0.16 (0.08)**
squared	**-0.09 (0.03)**	-0.01 (0.03)	**-0.09 (0.04)**	-0.01 (0.04)
7th / 4th	**-0.14 (0.07)**	-0.05 (0.06)	**-0.13 (0.08)**	0.03 (0.08)
squared	-0.01 (0.03)	-0.04 (0.03)	0.01 (0.03)	-0.03 (0.04)
8th / 5th	-0.05 (0.06)	**-0.15 (0.06)**	-0.06 (0.07)	-0.12 (0.07)
squared	0.01 (0.02)	-0.04 (0.02)	0.02 (0.03)	-0.04 (0.03)
Special Education				
6th / 3rd	0.03 (0.09)	0.05 (0.07)	0.04 (0.09)	0.06 (0.08)
7th / 4th	-0.08 (0.11)	-0.06 (0.08)	-0.09 (0.12)	-0.08 (0.09)
8th / 5th	0.06 (0.09)	0.10 (0.06)	0.06 (0.10)	0.12 (0.07)
Days Absent				
6th / 3rd	0.002 (0.005)	0.001 (0.005)	-0.002 (0.006)	-0.005 (0.006)
7th / 4th	0.004 (0.004)	0.001 (0.005)	0.005 (0.005)	0.001 (0.001)
8th / 5th	**0.012 (0.003)**	**0.012 (0.004)**	**0.012 (0.004)**	**0.018 (0.006)**
Days Suspended				
6th / 3rd	0.015 (0.013)	**0.125 (0.039)**	0.018 (0.016)	**0.152 (0.045)**
7th / 4th	0.006 (0.011)	0.014 (0.034)	0.001 (0.013)	0.019 (0.039)
8th / 5th	0.008 (0.009)	0.028 (0.027)	0.005 (0.011)	0.037 (0.031)
Ever Suspended				
6th / 3rd	**0.29 (0.08)**	**0.31 (0.12)**	**0.29 (0.10)**	0.22 (0.14)
7th / 4th	**0.39 (0.08)**	**0.45 (0.11)**	**0.42 (0.09)**	**0.40 (0.12)**
8th / 5th	**0.60 (0.07)**	**0.54 (0.09)**	**0.53 (0.09)**	**0.51 (0.11)**
Sample Size	20,858	22,657	10,439	11,344
Pseudo R-squared	0.218	0.185	0.189	0.179
X^2 (Test Scores)	163.12	158.07	114.24	130.75
X^2 (Behavior)	538.77	390.92	357.54	270.57
X^2 (Geography)	260.51	259.28	228.5	288.3

Notes: Each row gives the logit coefficient from a regression that predicts the probability that a student will ever be arrested as a function of the covariates listed above, plus dummy variables for missing test scores in each year and census tract-by neighborhood school fixed effects. The density of these arrest predictions is graphed in Figure III, and they are used to break students into the risk quintiles discussed in Section 3.1 The last 3 rows show test statistics for joint significance of the test score variables, the absence and suspension variables, and the geography fixed effects respectively. Values for missing data are imputed based on race and gender means, but only for students who were actually enrolled in CMS at the time. Coefficients in bold are sig. at the 5% level or greater.

Appendix Table IV: Effects of Winning the Lottery on Crime, by Race and Gender

	High School Sample				Middle School Sample			
	Male		Female		Male		Female	
	Black	Nonblack	Black	Nonblack	Black	Nonblack	Black	Nonblack
Felony Arrests	-0.148**	0.036	-0.043	-0.003	0.031	0.049	0.017	-0.023
	[0.064]	[0.047]	[0.037]	[0.003]	[0.091]	[0.051]	[0.024]	[0.017]
	{0.337}	{0.075}	{0.076}	{0.004}	{0.368}	{0.044}	{0.034}	{0.017}
Social Cost	-2,913**	375	-50	-20	-3,739**	489	-259	-44**
(murder trimmed)	[1,257]	[318]	[114]	[31]	[1,446]	[372]	[378]	[21]
	{5,399}	{607}	{336}	{44}	{5,887}	{570}	{727}	{50}
Sentence-Weighted	-7.41	2.91	-0.12	0	-9.91*	5.57	1.19	-0.16
	[6.10]	[2.14]	[0.52]	[0]	[5.71]	[4.19]	[1.83]	[0.12]
	{25.39}	{2.43}	{0.93}	{0}	{20.70}	{1.95}	{1.29}	{0.11}
Sample Size	610	404	559	318	649	432	797	442

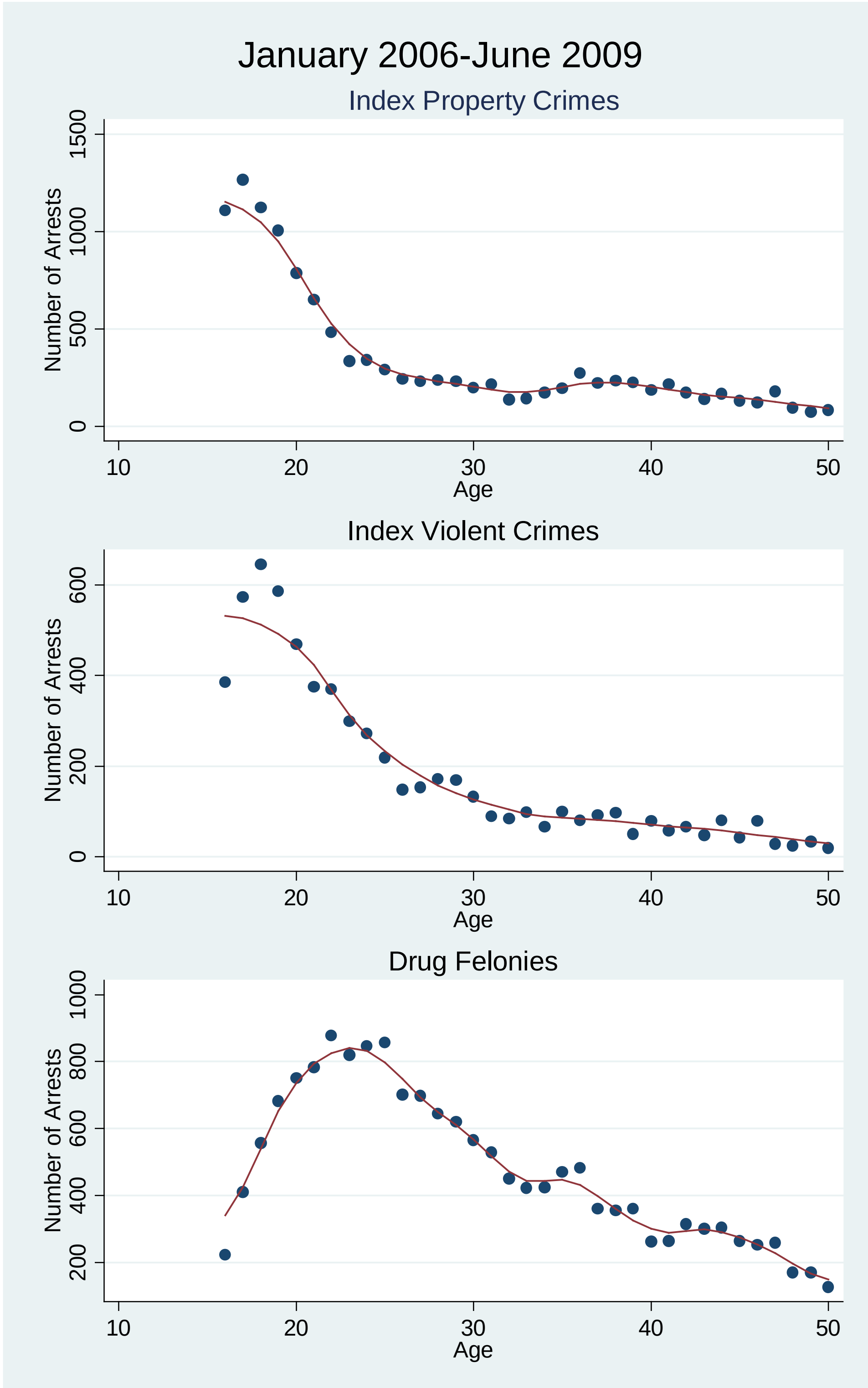
Notes : Each point estimate is from a regression like equation (1). The Xij vector includes free lunch status, prior math and reading scores, absences and out of school suspensions. Standard errors are below each estimate in brackets and are clustered at the lottery (i.e. choice by priority group) level. Control means are below the standard errors in curled brackets. * = sig. at 10% level; ** = sig. at 5% level; *** = sig. at 1% level.

Appendix Table V: Effect of Winning the Lottery on Crime - Alternate Specifications

Top Risk Quintile Only	High				Middle			
	OLS	Logit	Poisson	NBR	OLS	Logit	Poisson	NBR
Felony Arrests	-0.352**	-0.992***	-0.787***	-0.599***	0.101	0.226	0.020	0.069
	[0.126]	[0.317]	[0.243]	[0.228]	[0.180]	[0.405]	[0.268]	[0.236]
Total Days Incarcerated	-27.6	0.122	0.015	0.100	-38.3***	-0.39	-1.29***	-0.23
	[19.6]	[0.246]	[0.520]	[0.168]	[12.5]	[0.39]	[0.42]	[0.25]
Felony Charges								
Index Property	-0.239	-0.747	-0.697	-0.843*	0.261	0.648	0.430	0.286
	[0.250]	[0.539]	[0.544]	[0.477]	[0.173]	[0.565]	[0.328]	[0.399]
Index Violent	-0.089	0.384	-0.427	0.285	-0.376*	-0.690	-1.917**	-0.763*
	[0.199]	[0.719]	[0.878]	[0.595]	[0.201]	[0.457]	[0.773]	[0.453]
Drug Felonies	-0.342**	-1.680***	-1.454*	-0.996***	0.169	0.038	0.277	0.131
	[0.151]	[0.336]	[0.845]	[0.346]	[0.136]	[0.417]	[0.706]	[0.477]
Other Felonies	-0.287*	-0.708	-0.984	-0.285	-0.067	0.517	-0.336	0.091
	[0.145]	[0.702]	[0.668]	[0.619]	[0.123]	[0.361]	[0.412]	[0.350]
Sample Size	1014				1081			

Notes: Each estimate is from a regression like equation (2), where the lottery treatment is interacted with indicators for whether an applicant is in the 1st-4th or 5th arrest risk quintiles. The Xij vector includes only the predicted probability of arrest estimated in Section 3.1. Block bootstrapped standard errors (with lotteries as clusters) are below each estimate in brackets. The first column contains OLS estimates, repeating the results in Table 4. The second column estimates a logit and converts each outcome into an indicator variable. Columns 3 and 4 present results using poisson and negative binomial count models. Index Property Crimes are larceny, burglary and auto theft. Index violent crimes are murder, aggravated assault, robbery and rape. * = sig. at 10% level; ** = sig. at 5% level; *** = sig. at 1% level.

Online Appendix Figure I
Age Profile of Crime in Mecklenburg County



Notes: Includes all arrests, not just those matched to CMS students. The data begin at age 16, when youths are treated as adults by the criminal justice system in North Carolina.

Online Appendix Figure II