ONLINE APPENDIX

This online appendix includes supplementary material describing:

1. Boston SYEP Intervention and Experimental Design
2. Data Sources
3. Analysis Methods
4. Robustness checks
5. Cost Benefit Calculation
6. Additional References

1. Boston SYEP Intervention and Experimental Design

The Boston SYEP is administered by the Boston Mayor's Office of Workforce Development (OWD) and implemented by four non-profit community organizations, known as intermediaries. All Boston city residents aged 14 to 24 years are eligible for the program and youth apply directly to the program through one of the four intermediaries. The intermediaries are responsible for reviewing applications, supervising job placements, and delivering the program's career-readiness curriculum. Youth typically apply to the intermediary in their neighborhood. Administrative records indicate that less than 5 percent of youth apply to more than one agency.

Two of the intermediaries make use of random assignment to assign youth to jobs because of the high number of applications they receive for the limited number of SYEP jobs available. The analysis in this paper is restricted to youth who applied for a job for summer 2015 through Action for Boston Community Development (ABCD), one of the two intermediaries that make use of random assignment. The other intermediary that uses random assignment, the Department of Youth Employment and Engagement (DYEE), does so only on a partial basis where 60 percent of the jobs for a given employer are assigned randomly and the other 40 percent are selected. DYEE also chose not to implement the survey during the summer of 2015.

The enrollment period typically spans February through June, and applicants are notified of their lottery status and job assignment in late June. See Figure A1 for a timeline of the program and data collection. ABCD uses a computerized system with a random-assignment algorithm to select applicants based on their applicant ID numbers and the number of available slots as determined by the amount of funding ABCD receives each year. This system effectively assigns the offer to participate in the program at random, creating a control group of youth who apply to the SYEP but are not chosen. Of the 4,235 youth who applied to ABCD in 2015, a total of 1,186 were offered a job via random assignment (28 percent), leaving 3,049 individuals in the control group. Of those selected by the lottery, 83.6 percent accepted a job offer, with only a handful of youth dropping out of the program during the summer. As discussed in the main text, randomization successfully balanced all observable characteristics across treatment and control groups.

However, are the applicants served by ABCD representative of all youth age 14-24 years in the city of Boston? This question is important for demonstrating internal validity for the city of Boston and for city leaders seeking to bring the summer jobs program to scale. Figure A2 shows the distribution of youth age 0-17 based on data from the 2010 Census, with greater representation among Dorchester, Roxbury, and Mattapan. Table A1 shows that ABCD draws applicants from all 18 of the city's neighborhoods with a similar distribution of applicants showing greater representation among Dorchester (about 33 percent), Roxbury (about 10 percent), and Mattapan (about 9 percent). Applicants from other neighborhoods are also similarly represented in proportion to the distribution of youth as shown by the Census data: Hyde Park (about 6.6 percent), South Boston (about 6.4 percent), South End (about 6 percent), Roslindale (about 5.7 percent), Allston-Brighton (about 5 percent) and Jamaica Plain (about 4.5 percent).

Moreover, data from the 2011-15 5-Year American Community Survey indicate that ABCD applicants have similar gender and racial characteristics in comparison to the population of low-income Boston youth. Table A2 shows that although ABCD applicants are more likely to be younger, within that younger age group (age 14-17 years) the breakdown by gender and race is very similar. In general, it is reasonable to expect that youth applying to summer jobs programs would be younger given the greater difficulty that less experienced youth have in finding a job on their own.

2. Data Sources

2.1. Administrative Data on Criminal Justice Outcomes

The main outcome data consist of adult arraignment records from the Massachusetts Department of Criminal Justice Information Services and juvenile arraignment records from the Massachusetts Office of the Commissioner of Probation. This rich data source includes information on each criminal charge up through November of 2016—the date that the data were pulled by the two agencies. This includes the arraignment date, the seriousness of the crime (e.g., misdemeanor or felony), as well as the offense code that can be used to create categories for the type of crime (e.g. violent, property, drug, gun, and other).

Note that an arraignment is not the same as an arrest. After suspects are arrested and booked by police, they are arraigned as the first stage of court proceedings. During a typical arraignment, a person charged with a crime is called before a criminal court judge who reads the charges and asks if they need an attorney and how they plead. The judge sets bail and announces dates for future court proceedings

Administrative arraignment data avoid the limitations of self-reported crime like social desirability bias, which might be particularly problematic given that the treatment group may be less willing to admit wrongdoing than the control group to keep their job. Nonetheless, official arraignment records are not without limitations as measures of crime and violence. The criminal record data measures criminal activity only to the extent that an individual was arrested and "booked." It does not capture criminal activity that went undetected by police nor encounters with the police that did not result in official documentation.

In fact, police and prosecutors have the discretion of dropping the arrest without taking it to court, but this is different from arraignment which is simply being formally charged with the crime. It appears that there is little drop-off from arrest to arraignment, especially for juveniles who can only be held by police for up to six hours. During court hours, police must complete the booking process and then transport the juvenile to the Juvenile Court, according to the "2017-2018 Legal Issues Student Guide" published by the Municipal Police Training Committee [1].

Rather, it seems that there is more drop-off between arraignment and prosecution, which can be observed to some extent in the arraignment data, although not completely because not all arraignments have come to their final disposition during the post-program observation period. However, it appears that less than 20 percent of charges are dismissed or dropped. Similarly, a New York City report on stop-and-frisk found that between 2009 and 2012, almost one in six arrests (15.7%) were never prosecuted. [2] In Boston, the Suffolk County District Attorney claims that percentage is even lower, at least for homicide cases [3].

There are also differences in the arrest and arraignment rates for different types of crimes. For example, violent and property crimes are better-measured by arraignment data than other type of crimes such as non-violent drug and gun crimes where often there is no victim to report it. As of 2015, about 45 percent of violent crimes and 20 percent of property crimes were "cleared" through the arrest, charging, and referral of a suspect for prosecution. Differential arrest probabilities also help explain why the control means for drug, gun, and other arrests are much lower than violent crime and property crime arrests, despite nonviolent crimes being more common [4]. This could explain the ability to detect an impact on some types of crimes compared to others.

As a result, the arraignment data tend to understate the overall amount of crime, since many crimes do not result in an arraignment. In addition, arraignment data captures both criminal and police behavior at a point in time. However, the similarity of estimated program impacts across both administrative and self-reported crime data in another jobs-program evaluation, Job Corps, suggests that changes in police behavior or probability of being caught are unlikely to explain program effects. Moreover, because of the randomized design, the treatment and control groups are from similar neighborhoods and would be subject to the same policing behavior during the post-period.

Data were matched using name and date of birth where name was the one that used at the time of the individual's first court arraignment. The arraignment data therefore include all arraignments of an individual in the state of Massachusetts, even if he or she submits an alias at the time of arrest. There is little reason to believe that a summer jobs program would affect how names are recorded in the data, meaning that the matching error should be uncorrelated with treatment status. This is particularly true for youth applying through ABCD in Boston. There is a rigorous application process which requires verification of household income and receipt of public assistance for the purposes of being able to match youth to the appropriate funding streams that the organization must braid together each year across both government and charitable sources. As a result, the application process involves the signature of a parent to verify that the information is correct and to give consent for obtaining information from administrative schooling, employment, and criminal justice records. Thus, it is unlikely that name and date-of-birth would be differentially mis-reported across the treatment and control groups causing a better or worse match with the criminal justice records. In fact, similar proportions of youth were found to have been arraigned prior to the start of the program from the treatment group (4.1 percent) and the control group (3.6 percent).

To separate crimes by type, I categorize charges associated with each arraignment based on the offense code. Violent crimes include all crimes against a person: assault, homicide, sexual offenses, robbery, threats, kidnapping, and aggravated arson (arson when someone is known to be home). Property crime includes larceny, burglary, non-aggravated arson, motor vehicle theft, and vandalism. Drug crimes include both possession and dealing. Gun crimes include possession of a firearm, firearm violations, possession of ammunition, and carrying without a license. Other crimes include offenses such as possession of alcohol by a minor, operating under the influence, trespassing, disturbing the peace, cruelty to animals, and parole violations. Note that status offenses (or "child in need of assistance") as well as revocations (e.g. rules violations) were not included. I then count the number of pre- and post-program incidents of each type, defining "post" as occurring after the lottery notification date in June.

Note that the administrative criminal justice data are limited to arraignment within the state of Massachusetts. There are two types of criminal record checks in Massachusetts. The first type is a name-based criminal record check (CORI) that returns information on available Massachusetts arraignments. This type of criminal record check is done by submitting the name and date of birth for a person. That information is then searched against Massachusetts court records to determine if there is a possible record for that person. This type of criminal record check contains only Massachusetts information. The second type is a fingerprint-supported criminal record check. This type of check returns information on arrests made in Massachusetts as well as other states. This type of criminal record check is done by matching the fingerprints of a person against fingerprints collected by Massachusetts and other states. Since we do not collect fingerprints from youth, we are unable to perform this type of criminal record check.

Without a national database of arrests, it is difficult to assess the extent to which having state-level data is a limitation of the study. However, to bias the results it would have to be the case that treatment increases time spent outside the state and so reduces arrests without reducing criminal activity. However, all summer jobs were within the greater Boston area, so treatment did not directly encourage out-of-state travel. Thus, it seems implausible that differential censoring can explain the entire observed decrease in violent and property crimes.

2.2 Survey Data on Pre-/Post-Program Behavioral Outcomes

The survey was originally developed and implemented by the Youth Violence Prevention Collaborative, an initiative that began funding summer employment opportunities in Boston neighborhoods that had been identified by the Boston Police Department as having a high number of fatal and non-fatal shootings. Starting in the summer of 2012, the goal was to measure personal and social behaviors that correlate with youth violence and exposure to violence to determine whether summer employment could reduce the exposure of economically disadvantaged teens to risky, violent, and delinquent behaviors. This original survey was typically administered at the end of the summer to program participants and covered basic demographic information as well as questions on risky and delinquent behavior, community engagement, and general satisfaction with SYEP jobs and programming. During the summer of 2015, I worked with the Office of Workforce Development (OWD), I expanded the survey's content to include questions related to job readiness, post-secondary aspirations, and financial capability.

For most of the mediator variables of interest, improvement can only be measured simply as a 0/1 change for improvement. For example, in terms of academic aspirations an improvement is measured as switching from not wanting to attend college to wanting to attend college. Similarly, in terms of jobs readiness skills an improvement is measured as switching from not being able to write a resume to being able to write a resume.

However, for the social skill and community engagement questions it is possible to construct multiple measures. This is because these questions are measured using a Likert scale (e.g., Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree). I measure improvement in two ways. The main results in the paper measure improvement as any upward shift along the scale (e.g. switching from "Agree" to "Strongly Agree"). As a robustness check I also measure improvement as a "meaningful" positive change that is larger in magnitude and defined as switching from Disagree or Strongly Disagree to Agree or Strongly Agree. This definition has the advantage of capturing movements of at least two units (e.g., the smallest being Disagree to Agree) and is also well defined by not including Neutral as an option on either side. Table A13 in the online appendix shows that defining improvement using this more meaningful definition for the community engagement and social skill variables yields even stronger relationships between the short-term program impacts and the reduction in the number of arraignments.

In addition, OWD expanded the scope of the survey by engaging ABCD to conduct both a pre- and post-survey to measure changes over time for participants. The pre-survey was administered to participants during orientation in early July and the post-survey was administered in mid-August when participants pick up their last paycheck. Surveys were administered to participants on-site using a paper-based collection method. Although nearly the same number of individuals answered the pre- and post-surveys, these were not necessarily the same individuals as only 66.9 percent of individuals could be matched. However, testing for differential attrition between the pre- survey sample and the matched sample for both ABCD yields no statistically significant differences (see Table A4).

In addition, OWD also worked with ABCD to administer the post-survey to the control group to compare the experiences of participants to the counterfactual experiences of those who had applied but not been selected by the SYEP. The post-survey was administered to the control group on-line via email with a link to the survey web site using SurveyGizmo. The control group was offered the chance to win a free iPad mini for completing the survey. Yet despite several reminders and extensions, the response rates differed significantly across the treatment versus the control group. Indeed, although the number of respondents among the control group was similar (N=664), this represented a response rate of only 21.8 percent.

Moreover, although the control group was randomly selected, those who chose to respond to the post-survey were not. Unlike other household surveys, we know that the characteristics of the control group should be indistinguishable from those of the treatment group because the random assignment was shown to be balanced. This means that we can explore the sign of the bias by exploring how the observable characteristics differ between the two groups. Table A5 shows that relative to the treatment group, survey respondents from the control group exhibited characteristics that are on average associated with better economic, academic, and criminal justice outcomes. They were more likely to be older, female, identify as white or Asian, and indicate that they live in a two-parent household.

I argue that this bias goes *against* finding an impact for the Boston SYEP, given that the survey respondents in the control group exhibit demographic characteristics that would suggest a high bar for comparison. In the literature, each of the observable characteristics that differ for the control group relative to the treatment group has been shown to be associated with better employment, academic, and criminal justice outcomes (e.g., lower crime rates). In terms of employment, higher employment rates are observed among females, whites, and older youth [5]. In terms of academic outcomes, females are now more likely than males to attend college [6]. There is also a large literature explaining test score gaps that finds lower scores among African-American children and those living in single parent households [7]. In terms of criminal justice outcomes, age, male gender, and living in a single-parent home are significant predictors of re-offending among youth [8].

Moreover, youth in the control group who responded to the survey are likely to be more intrinsically motivated than those who did not. In general, surveying youth is difficult but particularly so when relying on email for deployment since youth are less likely than adults to use email for personal communication (e.g., texting friends is more common), especially during the summer when school is out. The control group was surveyed about their summer experiences via an email that came from the Boston Office of Workforce Development about a program for which they were not selected. As such, taking the time to open the email, read it, and complete the survey suggests a relatively high degree of motivation. One of the survey questions confirms this hypothesis: youth were asked why they wanted to work this summer. Among the respondents, youth in the control group were more likely than those in the treatment group to report wanting a summer job to learn more about college and less likely to report wanting to make money, have something to do, or stay out of trouble.

It is important to acknowledge the other limitations of self-reported survey data such as those raised in Meyer, Mok, and Sullivan (2015). In that paper, the authors measure the degree to which nationally representative surveys suffer not just from unit non-response but also from item non-response and measurement error by comparing survey results to administrative data. In terms of item non-response, this can be a problem, particularly when asking sensitive questions about behavior among developing youth. For example, one of the other intermediaries that works with court-involved youth (Youth Options Unlimited) chose to include a series of questions based on the Youth Behavioral Risk Survey that asked about risky behavior such as drug and alcohol use and physical violence. However, the non-response rate was too high (roughly 20 percent) so that these responses were not informative. In contrast, Table A6 shows that the item non-response rates for the survey questions used in the mediator analysis were less than 5 percent for both the ABCD treatment and control groups with no significant differences across the two groups.

Finally, in terms of measurement error, there is little room to assess the magnitude of this bias without access to administrative data that covers the same items as the survey. The only test for measurement error that I can perform is to compare the employment rate for the control group to what is found in the state quarterly wage and employment administrative data. Only 26.4 percent of those responding to the survey in the control group reported that they had worked during the summer. This rate is consistent with the quarterly wage record data provided by the Massachusetts Division of Unemployment Assistance, which shows that a similar proportion of youth in the control group (28.2 percent) reported working during the third quarter (July-September) of 2015. In addition, because I measure impact for the treatment group relative to control group, if we assume that the measurement error is random, then this would reduce efficiency but not cause bias. I do not have any reason to believe that measurement error would differ across the treatment and control groups.

3. Analysis Methods

To assess the impact of the Boston SYEP on criminal justice outcomes, I compare criminal records during the period following the intervention for youth offered SYEP placements (the treatment group) with the records for youth not offered placements (control group). I measure two outcomes of interest: whether an individual has been arraigned for any crime during the post-intervention period, and the number of arraignments per youth during the post-intervention period. Because SYEP participation is allocated via lottery, I obtain causal estimates using a simple comparison of means on the outcome of interest. This "Intent to Treat" (ITT) estimate measures the impact of *offering* the program on the outcome. In many cases, this is the policy relevant estimate because program administrators want to account for program take-up to assess the degree to which SYEP could reduce criminal activity among all the applicants, not just the participants. Nonetheless, because not all youth end up participating, the ITT will understate the effects of the program for those youth who choose to participate. To address this, I also provide estimates of treatment-on-the-treated (TOT).

3.1 Intent-to-Treat Analysis

Let Y_{it} denote a post-program outcome for individual i during post-randomization period t. I model this outcome as:

$$Y_{it} = SYEP_i\pi_1 + X_{i(t-1)}\beta_1 + \mu_{it1} \qquad (1)$$

where Y_{it} is the criminal justice outcome, $SYEP_i$ is a dummy variable indicating the individual received an offer to participate, $X_{i(t-1)}$ is a set of pre-existing baseline criminal justice outcomes and demographic characteristics, and μ_{it1} is a stochastic error term.

Although baseline characteristics are not necessary for identification, I include them in the regression to improve the precision of estimates by accounting for residual variation in the outcomes. Baseline covariates included in the main regressions are demographic characteristics collected during the application process including: age, gender, race/ethnicity, limited English, in school, public assistance, and homelessness. I also include baseline criminal justice outcomes captured by the administrative data during the pre-program period. To determine the number of baseline arraignments, pre-program is defined as the 17 months prior to random assignment (February 2014 through June 2015). To determine if an individual has ever been charged with a crime, the pre-program period is defined as any time prior to July 2015. None of the substantive conclusions are different if these variables are excluded from the outcome regressions, but the covariates do improve precision.

3.2 Treatment-on-the-Treated Analysis

Nonetheless, because not all youth end up participating, the ITT will understate the effects of the program for those youth who choose to participate. Under the usual relevance and exogeneity assumptions for instrumental variables, this latter set of effects can be recovered from the experimental data.[1] I perform this estimation through a two-stage least squares strategy, in which random assignment ($SYEP_i$) is an instrument for actual participation (P_{it}), and P'_{it} is the predicted probability of participation from equation (2):

$$P_{it} = SYEP_i\, \pi_2 + X_{i(t-1)}\, \beta_2 + \mu_{it2} \tag{2}$$

$$Y_{it} = P'_i\, \pi_3 + X_{i(t-1)}\, \beta_3 + \mu_{it3} \tag{3}$$

If all youth respond the same way to the program (i.e., if treatment effects are constant across youth), then equations (2) and (3) also yield an estimate of the average treatment effect (ATE) across this population of disadvantaged youth. Given that treatment effects are likely to be heterogeneous across youth, then the coefficient π3 estimates a local average treatment effect—the effect of participation on those who comply with random assignment.[2] Because there is no control crossover (no always-takers) in this setting, π3 provides an estimate of the treatment-on-the-treated.

3.3 Functional Form

While ordinary least squares provides the best linear unbiased estimate of the treatment effect under the Gauss-Markov assumptions, I also explore the robustness of the results to alternative assumptions. Specifically, I relax the linear functional form assumption by using non-linear specifications. To analyze treatment-control differences in the number of arrests – a count variable – I use a Poisson quasi-maximum likelihood estimator (QMLE). The consistency of this estimator only requires the correct specification of the conditional mean, not the entire distribution [9]. I also use Huber-White robust standard errors to allow for over-dispersion, relaxing the Poisson distributional constraint that the mean equals the variance. To analyze differences in the likelihood of being arrested, a 0/1 dependent variable, I use a probit estimator.

3.4 Exploration of Program Mechanisms

---

[1] For the random assignment variable, $SYEP_i$, to be a valid instrument, it must be correlated with program participation, P_{it}, and uncorrelated with μ_{it3}.
[2] When treatment effects are heterogeneous, $SYEP_i$, must also satisfy a monotonicity condition to be a valid instrument. In particular, random assignment must make everyone weakly more likely to participate and no one less likely.

Ideally, a full mediation analysis would be used to generate evidence for how the SYEP program achieved its effects using measures of the mediating variable as well as the dependent and independent variable [10, 11]. The first step is to estimate a significant relationship between the dependent variable of interest (Y_{it}) and the independent variable ($SYEP_i$) using equation (1) above.

Second, a significant relationship is estimated between the hypothesized mediating variable (M_{it}) and the independent variable ($\mathrm{SYEP_i}$) using the following equation:

$$M_{it} = SYEP_{it}\,\pi_4 + X_{it}\,\beta_4 + \mu_{it4} \tag{4}$$

where M_{it} is one of the short-term program outcomes (e.g., social skills), $\mathrm{SYEP_i}$ is a dummy variable indicating the individual received an offer to participate, and X_{it} is a set of demographic characteristics collected at the time of the survey.

Third, the mediating variable (M_{it}) is shown to be significantly related to the dependent variable (Y_{it}) when both the independent variable and mediating variable are included as predictors:

$$Y_{it} = SYEP_i\,\pi_5 + X_{i(t-1)}\,\beta_5 + M_{it}\,\gamma + \mu_{it5} \tag{5}$$

To be a valid mediator, the coefficient π_3 relating the independent variable to the dependent variable in equation (5) must be smaller (in absolute value) than the coefficient π_1 relating the independent variable to the dependent variable in the equation (1) without the mediating variable. Researchers often test whether there is complete or partial mediation by testing whether $\pi 5$ is statistically significant, which is a test of whether the association between the independent and dependent variable is completely accounted for by the mediator.

Due to data limitations, I am unable to undertake the typical mediation analysis described above. This is because the post-survey was administered to the control group anonymously, rather than confidentially as was done for the treatment group. As such, I can only link the self-reported survey responses to the longer-term criminal justice outcomes for youth in the treatment group who responded to the survey. Nevertheless, it is still possible to explore whether improvement in the short-term behavioral impacts are correlated with subsequent improvement in the criminal justice outcomes to shed light on the program's mechanisms. I do this in three ways.

First, I modify equation (5) as follows:

$$Y_{it} = SYEP_i\,\pi_6 + X_{i(t-1)}\,\beta_6 + \Delta M_i\,\delta + \mu_{it6} \tag{6}$$

On the left-hand side, the dependent variable is one of the longer-term criminal justice outcomes (e.g., number of crimes per youth) while on the right-hand side is a dummy indicating positive improvement for a specific short-term program impact ΔM_i (e.g., ability to resolve conflicts with a peer). A negative and significant coefficient on ΔM_i indicates that improvement in the short-term program impact observed during the summer of participation is negatively correlated with longer-term criminal behavior. Moreover, if the coefficient on the $SYEP_i$ dummy in equation (6) is smaller in magnitude than that in equation (1), this would suggest that ΔM_i plays a role in achieving the longer-term impact separate from simply being assigned to treatment.

However, only youth in the treatment group who participated will have responded to the survey. As such, it is still possible that the observed changes in the short-term program measures from the survey are correlated with other unobserved factors (e.g. motivation to participate) that are driving the longer-term reduction in criminal behavior. I address this in two ways. First, I use a two-stage least squares to estimate the impact of the short-term behavioral impacts on the longer-term criminal justice outcomes using the SYEP treatment dummy as an instrument for participation and include ΔM_i as a control:

$$P_{it} = SYEP_i\, \pi_6 + X_{i(t-1)}\, \beta_6 + \Delta M_i\, \zeta + \mu_{it6} \tag{7}$$

$$Y_{it} = P'_i\, \pi_7 + X_{i(t-1)}\, \beta_7 + \Delta M_i\, \zeta + \mu_{it7} \tag{8}$$

Again, if the coefficient on ΔM_i is negative and significant and the coefficient on the $SYEP_i$ dummy is smaller in magnitude than that in equation (1), this would suggest that ΔM_i is a potential mediator.

Second, I use an alternative specification for the mediator analysis to test whether these same short-term program measures are driving the reduction in crime among only program participants completing both surveys (see Table A14) using the equation (9):

$$Y_{it} = X_{i(t-1)}\, \beta_4 + \Delta M_i\, \zeta + \mu_{it} \tag{9}$$

Note that the mediator analysis implicitly assumes that there was no change in the short-term program measures for youth in the control group. I argue that this assumption is plausible if the analysis is restricted to those short-term program impacts for which there was both significant improvement over time among participants and for which the gains were significant relative to the control group by the end of the summer.

In fact, there is an entire literature on the **loss** of skills among youth over the summer, particularly among disadvantaged groups. A meta-analysis summarizing the findings from the literature regarding academic skills concluded that concluded that: (1) on average, students' achievement scores declined over summer vacation by one month's worth of school-year learning, (2) declines were sharper for math than for reading, and (3) the extent of loss was larger at higher grade levels [12].

Moreover, summer learning loss is not limited to academic skills but is likely to also affect ***social*** skills. This is because although summer has the potential for more social interaction, youth report being lonelier. A recent survey of 2,000 youth found that more than half of teenagers feel isolated during their time off from school, with a quarter saying the summer is their loneliest time of the entire year, rising to 29 per cent among girls [13]. While changes in technology have made staying in touch with friends easier than ever, an over-reliance on smartphones and social media apps appears to have left teenagers feeling more isolated as they substitute quality time with friends for texting. The study found that although almost two thirds of teens will talk to their friends every day on social media during the summer, only 14 percent will see them face-to-face.

In addition, there is a newer literature regarding "summer melt"—a surprisingly common scenario in which high-school graduates apply, are accepted, and say they plan to enroll in college—but don't. This literature supports the assumption that academic aspirations to attend college do not typically increase over the summer, even among those who have been accepted. For example, 22 percent of the lowest income, college-intending students in Boston uAspire failed to matriculate into college in the fall after high school graduation, compared to an 18 percent among all other students [14]. A study reporting on two randomized trials finds that offering college-intending graduates two to three hours of summer support and coaching increased college enrollment by 3 percentage points overall, and by 8 to 12 percentage points among low-income students [15].

4. Robustness Checks

4.1 Treatment-on-the-Treated results

As discussed above in the analysis methods section, because not all youth end up participating, the ITT estimates will understate the effects of participating in the program for those youth who choose to participate. Table A7 reports the Treatment-on-the-Treated (TOT) estimates that show the effect of the Boston SYEP for those who choose to participate. For that group, violent crime falls by -3.6 arraignments per youth and property crime decreases by -2.9 arraignments per youth. The TOT estimates are only slightly larger than the ITT estimates, likely because the take-up rate is so high (about 85 percent).

4.2 Decrease in number of arraignments per youth but no decrease in share of youth arraigned

Similar to Heller (2014), Figure 2 in the paper shows that the Boston SYEP has a significant impact on reducing the frequency of criminal arraignments among youth but not the likelihood of ever being arraigned. Table A8 shows that the ITT estimate of the program's impact on reducing the likelihood of ever being arraigned during the post-program period is both economically and statistically insignificant. Only the reduction in ever being arraigned for a gun crime is marginally significant using OLS and this result loses significance when using a probit model which is more appropriate for a 0/1 outcome such as ever having been arraigned during the 17-month post-program period. The program's lack of impact on the likelihood of ever being arraigned holds even if we adjust the post-program observation window. Table A9 shows that even using a more restricted observation window from month six forward (the period over which we see the cumulative reduction in arraignments become significant) yields no economically or statistically significant reduction in ever being arraigned.

Although some might conclude that because the program does not reduce the likelihood of ever being arraigned, only the frequency of arraignments, SYEP are not effective, most criminologists would disagree. First, the reduction in the number of crimes is an economically meaningful impact—particularly for violent and property crimes that are costly to both individuals and society (see Table A16 below for cost benefit calculations).

Second, these findings are consistent with prior research in criminology regarding the likelihood of youth to participate in delinquent and criminal activities. The likelihood of ever committing a crime increases with age as part of the normal course of adolescence through age 25 when delinquency and criminal activity naturally decrease due to maturity without any intervention [16]. Indeed, Table A10 shows that the relative reduction in the number of arraignments for violent and property crimes was driven by a smaller increase over time among the treatment group. Yet the share of youth being arraigned for any crime increased for both the treatment and control groups.

Third, one of the potential rationales from the criminology literature through which SYEP reduce crime is by "disrupting routine activities" that provide likely offenders with suitable targets and a lack of supervision or guardianship [17]. Thus, it is likely that SYEP do not completely incapacitate but rather disrupt the frequency with which such youth engage in crime. This rationale is supported by focus group discussions with SYEP participants who had been court-involved prior to the program. When asked whether the program reduced the opportunity to engage in crime, the respondents acknowledged that they avoided interacting with other gang-related youth during the week because they had to get up early the next day for work or else they would be fired.

Finally, SYEP do not change the household situation, neighborhood, or school environment of participating youth—contextual factors that are also important in explaining the likelihood of participating in delinquent and criminal activity. It is likely that for a subset of youth, these other contextual factors would make them predisposed to crime. This is consistent with the work of Heller and Davis (2017b) regarding the considerable heterogeneity in program outcomes across groups. In that paper, positive employment outcomes are observed among youth with no criminal history. Conversely, it is likely that the reduction in crime is observed among marginal or "at-risk" youth that would be inclined to engage in criminal activity in the first place.

In fact, the aggregate effect is entirely driven by reducing the number of arraignments among youth ever arraigned during the post-period. Table A11 shows the distributions of the number of arraignments for those arraigned at least once during the 17-month post-program window. The unadjusted means as well as the OLS and Poisson estimates show that among those ever arraigned during the post period, there was a significant reduction in the ***total*** number of arraignments per youth for the treatment group compared to the control group (-1.3 fewer arraignments). This reduction in the total number of arraignments was driven by a significant reduction in violent and property arraignments for both misdemeanors and felonies.

Moreover, it does not appear that the program reduces the overall likelihood of recidivism, but rather the number of post-program arraignments among those who had ever been arraigned prior to the program. Table A12 shows that while the re-arraignment rate is similar for both the treatment and control groups (about 35 percent), the number of post-program arraignments per youth for violent and property crimes is significantly lower for the treatment group relative to the control group. In comparison, there is no economically or statistically significant difference in the number of post-program arraignments per youth among those never arraigned prior to the program. These results are also consistent with those from one of the other program intermediaries, Youth Options Unlimited (YOU), that works exclusively with gang-related youth referred to them by the Boston Police Department. Although we can only compare changes over time as there is no control group for YOU, Table A15 shows that the ***total*** number of arraignments decreased significantly among participants (-3.5 per youth) including a significant decrease in violent crime arraignments (-1.5 per youth), split evenly among misdemeanors and felonies.

4.3 Heterogeneity in reducing the number of arraignments overall versus by type of crime

Despite there being no significant difference in the overall number of arraignments per youth, violent-crime arraignments among the treatment group were 35 percent lower relative to the control group, with roughly 3.1 fewer arraignments per 100 youth. A similar impact was found for property crimes (-2.2 fewer arraignments per 100 youth or -29 percent). There were no significant changes in arraignments for the other types of crimes (gun, drug, or other). These results are very similar to those of Heller (2014) who finds that the Chicago SYEP reduced violent crime arrests but did not reduce the number of crimes overall or for other categories of crime.

One reason for finding a significant impact on violent and property crime arraignments but no other types of arraignments could be due to violent crime and property crime being better-measured outcomes in the administrative data. As of 2015, about 20 percent of property crimes and 45 percent of violent crimes were "cleared" through the arrest, charging and referral of a suspect for prosecution. Differential arrest probabilities also help explain why the control means for drug, gun, and other arrests are much lower than violent crime and property crime arrests, despite nonviolent crimes being more common [4].

Indeed, there was a slight but not statistically significant uptick in drug crimes and other crimes such as disturbing the peace or interrupting a school assembly which may account for why the decline in the overall number of arraignments is not statistically significant. For the former, it could be the case the additional income from working is spent on crime-inducing goods such as drugs. For the latter, it may be the case that incidents such as disturbing the peace continue to occur as frequently as before but no longer escalate into violent crimes. Interestingly, similar reductions in arraignments were observed regardless of the seriousness of the crime (i.e., misdemeanor versus felony).

4.4 Subgroup Analysis

Of course, SYEPs do not change the household, neighborhood, or school environment of participating youth—contextual factors that also are important in explaining criminal activity. It is likely that for a subset of youth, such contextual factors would make them predisposed to engage in criminal activity. For example, males typically offend at a higher rate than females. Similarly, youth tend to be more likely to commit crimes as they age because they have less supervision and more opportunity—especially if they can drive and are no longer in school. Finally, youth with less advantageous socioeconomic characteristics, such as those living in poverty or who are homeless, have been shown to be more likely to engage in crime [19, 20]. Finally, youth who have previously been arrested are at higher risk for being arrested in the future [21].

As such, it is natural to ask whether SYEPs might have a disproportionate effect on subgroups that are more likely to engage in crime. It should be noted that these subgroup analyses were not pre-specified, but rather are exploratory, and are subject to the usual bias arising from multiple hypothesis testing. Still exploratory subgroup analyses can be useful for generating new hypotheses and for robustness checking.

There are two primary ways to adjust for multiple hypothesis testing. The first is the False Discovery Rate (FDR) which controls for the expected proportion of false rejections below α or the probability that a null rejection is a Type I error. This increases the power of individual hypothesis tests in exchange for allowing some specified proportion of rejections to be false. I implemented the FDR control using code provided by Michael Anderson (Berkeley) which uses a procedure proposed by Benjamini and Hochberg [22,23]. Using this method, I find that while there are positive and significant impacts for three out of the four subgroups using the unadjusted p-values, the results are only significant for males when controlling FDR at q=0.10.

The second method is the Familywise Error Rate (FWER) which is more conservative than the FDR and controls the probability than at least one hypothesis is falsely rejected below the level α. To implement the FWER control, I used the "wyoung" stata command written by Julian Reif (University of Illinois) which adjusts the p-values using the free-step down resampling method [24]. This method relies on a bootstrap resampling technique that simulates data under the null hypothesis. It calculates the probability of observing results as extreme as the ones estimated given the correlational structure and underlying distributions of the data, and the number of tests run within each family of outcomes [25]. Using this more conservative approach, the adjusted p-value for the treatment impact on violent-crime arrests for males is p = 0.073. Although this is slightly above a 95 percent cutoff, it is still statistically significant at the 10 percent level and suggests the decrease is unlikely to be due to chance.

Overall, only the decrease in violence for males is robust to adjustments for multiple hypothesis testing. I report both the adjusted p-values from the FWER and the q-values from the FDR respectively in Table 4 for males, youth age 18-24 years, those receiving public assistance, homeless youth, and those who had ever been arraigned prior to participating in the summer jobs program. Among the five subgroups, males are the only sub-group for which there is consistent evidence that the Boston SYEP has a greater impact on reducing arraignments for violent (-7.1 arraignments per 100 youth) crimes. Although I find some evidence that the program has a greater impact on reducing property crime among older youth and violent crime among homeless youth and those who had been arraigned prior to the program, these estimates do not remain statistically significant when adjusting the p-values to account for multiple hypothesis testing using the family-wise error rate (FWER).

4.5 Comparison of results to other SYEP studies

Three prior studies have examined the impact of SYEP on crime in different cities. The first studied the impact of the New York City SYEP on crime but using qualitatively different outcomes (e.g., incarceration and mortality) measured by state prison and death records (Gelber, Isen, and Kessler, 2014). The second is a more similar study that uses arrest data to estimate the impact of the Chicago SYEP by type of crime (Heller 2014). The third study reports program impacts on violent crime arrests from the Chicago program for subgroups predicted to have a positive impact based on a large set of observable characteristics yet does not detect heterogeneity in violence (Davis and Heller 2017b).

Comparing the results from Boston to those from Chicago for Heller's 2014 study shows that the impacts are quite similar, confirming that the positive impacts from Chicago are indeed attributable to SYEP rather than a city-specific effect. Figure A3 shows that across the two cities the mean number of post-program arraignments for violent crimes is almost identical for both the control and treatment groups and the ITT estimate is very similar (-3.13 crimes per 100 youth in Boston compared to -3.95 per 100 youth in Chicago).

In terms of property crime, while the post-program treatment means are very similar across the two programs, the control means are quite different, resulting in a significant reduction of -2.21 crimes per 100 youth in Boston but no significant reduction in Chicago. One reason for the difference could be that Heller (2014) does not include vandalism in the property crime category. Another reason could be that the Chicago program was specifically designed primarily as a violence-reduction intervention. Program operators focused on recruiting a population of youth at high risk of violence involvement in the city, using individual-level administrative data from multiple city agencies to capture factors such as previous justice system and gang involvement, as well as truancy and school engagement.

Finally, although the post-program means for drug crimes in Boston (about 0.01) are much lower than those in Chicago (about 0.05), both studies find no significant impact on drug crimes.

Perhaps more striking is the similarity in the cumulative reduction in the number of arraignments for violent crimes across the two studies. Figure A4 shows that there is a slight but insignificant uptick in crimes for the treatment group relative to the control group during the program, again suggesting that the program does not completely "incapacitate" youth during the summer. However, in both cases the cumulative reduction in violent crime becomes significant six to seven months after the program's start and then continues to fall during the observation period through months 16 (Chicago) or 17 (Boston).

Finally, this paper is also comparable to Davis and Heller (2017a) which investigates the heterogeneity in outcomes across subgroups as well as potential mechanisms. The Boston results differ in terms of the former but are consistent with the latter. For example, Davis and Heller (2017) do not detect subgroup differences for the number of violent crimes using traditional methods whereas the Boston program shows greater impacts for reducing violent crime among males. This difference could be due to the Chicago study having a smaller sample size and being block randomized at the school-gender level.

In terms of mechanisms, Davis and Heller (2017b) find that employment benefiters commit more property crime than their control counterparts, yet non-benefiters show a decline in violent crime. The authors conclude that their results do not seem consistent with the typical SYEP rationale that improved human capital and better labor market opportunities create a higher opportunity cost of crime. They discuss several alternative mechanisms and conclude that brief youth employment programs can generate substantively important behavioral change, but for different outcomes, different youth, and different reasons than those most often considered in the literature.

The results from the Boston program regarding the behavioral impacts are consistent with Davis and Heller (2017b). The mediation analysis shows that changes in social skills are correlated with reductions in crime among the participants, but changes in other short-term outcomes such as job readiness and academic aspirations are not. Like Davis and Heller (2017b), this evidence goes against the theory that SYEP reduce crime by improving human capital or increasing the opportunity cost of crime.

Other studies have found that SYEP participation is associated with short-term gains in academic and employment outcomes but that these impacts do not persist into the long-term. For example, Schwartz et al. (2015) found small but significant increases in the share of New York City SYEP participants taking and passing statewide high school exams relative to the control group. A related study demonstrated significant increases of one to two percent in the treatment group's school attendance during the year following participation, with larger improvements for students aged 16 years and older with prior low baseline attendance (Leos-Urbel, 2014). However, other research has indicated that the New York City program did not have a positive effect on high school graduation (Valentine et al., 2017) or college enrollment (Gelber et al., 2014). Two studies found that the New York City SYEP caused average earnings and the probability of employment to increase during the program but that these effects subsequently faded (Gelber et al. 2014, Valentine et al. 2017).

In comparison, the short-term impacts on job readiness and college aspirations associated with the Boston program do not necessarily contradict the previous evidence in the literature regarding longer-term outcomes on employment and college-going. For example, improvements in college aspirations may be constrained by other factors that affect enrollment such as financial resources or subsequent summer melt. In addition, youth may gain job readiness skills in the summer but because they are in school during the following year, these short-term outcomes may not translate into longer-run improvement in employment if they choose not to work while in school. The MDRC evaluation of the NYC program is consistent with this finding and shows that the treatment group was 54 percentage points more likely to work during the summer compared to the control group. Small but statistically significant impacts on employment were also detected in years 1 and 2 that then faded during subsequent years.

4.6 Alternative measures and specifications for mediator analysis

As discussed above in section 2, due to data limitations, I am unable to undertake the typical mediation analysis because the post-survey was administered to the control group anonymously, rather than confidentially as was done for the treatment group. As such, I can only link the survey responses to the longer-term criminal justice outcomes for youth in the treatment group who responded to the survey. Nevertheless, it is still possible to explore whether improvement in the short-term behavioral impacts are correlated with subsequent improvement in the criminal justice outcomes to shed light on the program's mechanisms using the following equation:

$$Y_{it} = SYEP_i\, \pi_6 + X_{i(t-1)}\, \beta_6 + \Delta M_i\, \delta + \mu_{it6} \tag{6}$$

On the left-hand side, the dependent variable is one of the longer-term criminal justice outcomes (e.g., number of crimes per youth) while on the right-hand side is a dummy indicating positive improvement for a specific short-term program impact ΔMi (e.g., ability to resolve conflicts with a peer). A negative and significant coefficient on ΔMi indicates that improvement in the short-term program impact observed during the summer of participation is negatively correlated with longer-term criminal behavior. Moreover, if the coefficient on the SYEPi dummy in equation (6) is smaller in magnitude than that in equation (1), this would suggest that ΔMi plays a role in achieving the longer-term impact separate from simply being assigned to treatment.

For most of the mediator variables, improvement can only be measured simply as a 0/1 change for improvement. For example, in terms of academic aspirations an improvement is measured as switching from not wanting to attend college to wanting to attend college. Similarly, in terms of jobs readiness skills an improvement is measured as switching from not being able to write a resume to being able to write a resume.

However, for the social skill and community engagement questions it is possible to construct multiple measures. This is because these questions are measured using a Likert scale (e.g., Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree). For example, we can imagine measuring improvement in two ways. In the paper, I measure this as any positive change (e.g. switching from Strongly Disagree to Disagree or any upward shift). Alternatively, we could also measure this as a "meaningful" positive change that is larger in magnitude.

As an alternative approach, I defined a "meaningful" improvement as switching from Disagree or Strongly Disagree to Agree or Strongly Agree. This definition has the advantage of capturing movements of at least two units (e.g., the smallest being Disagree to Agree) and is also well defined by not including Neutral as an option on either side. Table A13 shows that defining improvement using this more meaningful definition for the community engagement and social skill variables yields even stronger relationships between the short-term program impacts and the reduction in the number of arraignments. Even when controlling for the SYEP treatment variable as well as demographic covariates, improvements in social skills are negatively correlated with the reduction in violent and property crime arraignments and the magnitude of the impact is larger than the specifications using just any improvement. In addition, the impact of the community engagement measures becomes statistically significant when using these more meaningful measures. These findings support the "dosage" hypothesis that greater changes in short-term behavioral outcomes produce greater reductions in crime.

Finally, Table A15 shows that the relationship between improvements in short-term measures of social skills and reductions in criminal activity is even stronger when restricting the analysis to just the participants. Improvements in managing emotions and resolving conflict with a peer are correlated with a reduction in both violent and property crime of 2 to 5 arraignments per 100 youth. In addition, improvements in asking for help is associated with a decrease of 2.5 arraignments per youth for property crimes.

5. Cost Benefit Calculation

A key question from a policy perspective is whether the benefits to society from the program outweigh the program's costs. Although it is somewhat premature to perform a full cost benefit analysis until other key outcomes related to schooling and employment have been measured, I provide some back-of-the-envelope calculations comparing the short-term benefits from the reduction in crime to the program's costs.

The cost of administering the program for the City of Boston was about $2,000 per participant, which includes an average of just over $1,400 in wages. From a societal perspective, the wage cost is simply a transfer from the government to the youth and so is not generally counted as a net change in overall resources. This leaves an administrative program cost of $600, although if one wanted to separate the costs and benefits that accrue to the government, participants, and society, then wages would appear as a cost to the government and a benefit to participants. Note that this is the budgetary cost to the City for funding the program. It may understate the costs from a broader perspective, as it does not include the opportunity cost of city staff, time donated by program providers, or the deadweight loss involved in raising the tax dollars.

Valuing the benefits of reduced violence requires assigning a monetary value to each type of crime. I use estimates of the social costs of crime (tangible losses plus quality of life) and inflate them to 2015 dollars using the CPI [18]. Because there are only two murders in the data (in the control group) and the cost is so high (about $5m in 2015 dollars), it skews the data and inflates the standard errors. To address this, I “trim” the violent crime data by substituting the cost of the next highest costly crime (sexual assault) for murder which provides just enough power to detect an effect.

Although there is a great deal of uncertainty in these estimates, reasonable specifications suggest that the program’s benefits may already outweigh the costs. Table A16 shows that the estimated reduction in costs is -$1,793 for violent crimes and -$135 for property crimes for a combined total cost savings of -$1,928. This benefit to victims certainly outweighs the program costs of $600 per participant. Moreover, it is likely to be an underestimate since these calculations are based on administrative data which captures only about half of violent incidents and excludes the cost to the criminal justice system of arresting, trying, and potentially incarcerating the offender.

6. Additional References

[1] Municipal Police Training Committee (https://www.mass.gov/files/documents/2017/10/31/FINAL%202017-2018-STUDENT.pdf).

[2] New York State Office of the Attorney General. 2013. A Report on Arrests Arising from New York City's Stop-and-Frisk Practices. (https://www.nlg-npap.org/sites/default/files/OAG_REPORT_ON_SQF_PRACTICES_NOV_2013.pdf)

[3] Boston Suffolk County District Attorney (http://www.suffolkdistrictattorney.com/clearing-up-clearance-and-conviction-rates/).

[4] Gramlich, John. 2017. "Most violent and property crimes in the U.S. go unsolved." Pew Research Center, http://www.pewresearch.org/fact-tank/2017/03/01/most-violent-and-property-crimes-in-the-u-s-go-unsolved/)."

[5] Child Trends. 2017. Key Facts about Youth Employment. Retrieved from: https://www.childtrends.org/indicators/youth-employment

[6] Hugo Lopez, M. and Gonzalez-Barrera, A. 2014. "Women's college enrollment gains leave men behind." Pew Research Center. http://www.pewresearch.org/fact-tank/2014/03/06/womens-college-enrollment-gains-leave-men-behind/

[7] Jencks, C. and Phillips, M. 1998. "The Black-White Test Score Gap: Why It Persists and What Can Be Done." The Brookings Institution. https://www.brookings.edu/articles/the-black-white-test-score-gap-why-it-persists-and-what-can-be-done/

[8] Cottle, C., Lee, R. & Heilburn, K. 2001. The Prediction of Criminal Recidivism in Juveniles A Meta-Analysis. Criminal Justice and Behavior. 28(3), 367-94.

[9] Wooldridge JM., in Handbook of Applied Econometrics Volume II: Microeconomics, M. H. Pesaran, P. Schmidt, Eds. (Blackwell, Oxford, UK, 1997), pp. 352–406.).

[10] Baron RM & Kenny DA. 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Personal Soc Psychology, 51, 1173–82.

[11] Keele, L., Tingley, D. & Yamamoto, T. (2015). Identifying Mechanisms Behind Policy Interventions Via Causal Medication Analysis. Journal of Policy Analysis and Management, 34, 937–963.

[12] Cooper H., Nye B., Charlton K., Lindsay J., Greathouse S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. Review of Educational Research, 66(3), 227–268.

[13] Panayiotou S., Newton S., Matthews P., Webster H., Andersson D. 2017. National Citizen Service 2016: Evaluation Technical report. Kantar Public, December.

[14] Castleman, B. and Page, L. A Trickle or a Torrent? Understanding the Extent of Summer 'Melt' among College-Intending High School Graduates. Social Science Quarterly, 95(1).

[15] Castleman, B., Page, L., Schooley, K. 2014. "The Forgotten Summer: Does the Offer of College Counseling after High School Mitigate Summer Melt among College-Intending, Low-Income High School Graduates?" Journal of Policy Analysis and Management, 33(2), 320-44.

[16] Monahan, K., Steinberg, L., & Piquero, A.R. (2015). Juvenile justice policy and practice: A developmental perspective. Crime & Justice, 44, 577−619.

[17] Cohen, L.E. & Felson, M. (1979, August). Social change and crime rate trends: A routine activity approach. American Sociological Review, 44, 588–608.

[18] Miller T.R., Cohen M.A., Wiersema B. 1996. "Victim costs and consequences: A new look," National Institute of Justice Research Report (U.S. Department of Justice, National Institute of Justice).

[19] HUD, Office of Policy Development and Research. 2106. Neighborhoods and Violent Crime. Evidence Matters. Summer.

[20] Baron, SW. & Hartnagel TF. (1998). Street youth and criminal violence. Journal of Research in Crime and Delinquency, 35.

[21] Cottle, C., Lee, R. & Heilburn, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. Criminal Justice and Behavior. 28, 367-94.

[22] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc., B **57**, 289–300 (1995).

[23] M. L. Anderson, Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. J. Am. Stat. Assoc. **103**, 1481–1495 (2008). doi:10.1198/016214508000000841

[24] P. H. Westfall, S. S. Young, Resampling-Based Multiple Testing: Examples and Methods for P-value Adjustment (Wiley-Interscience, New York, 1993).

[25] Jones, D., D. Molitor, and J. Reif. 2018. "What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study." National Bureau of Economic Research Working Paper No. 24229.

Figure A1. Timeline of Boston SYEP Program and Data Collection

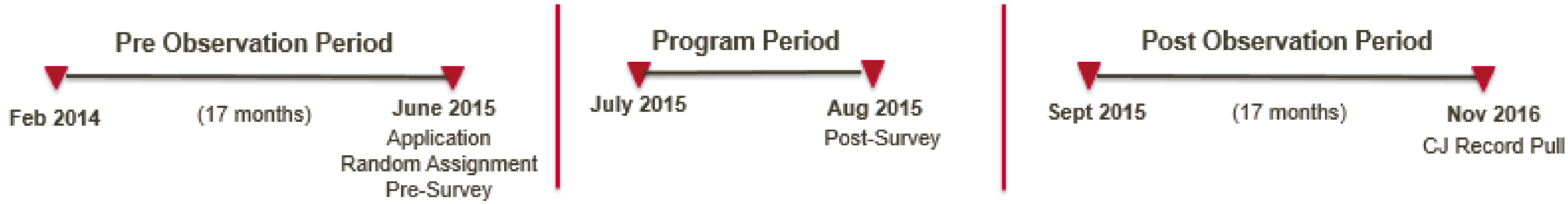

Figure A2. Distribution of Boston Youth Population by Neighborhood.

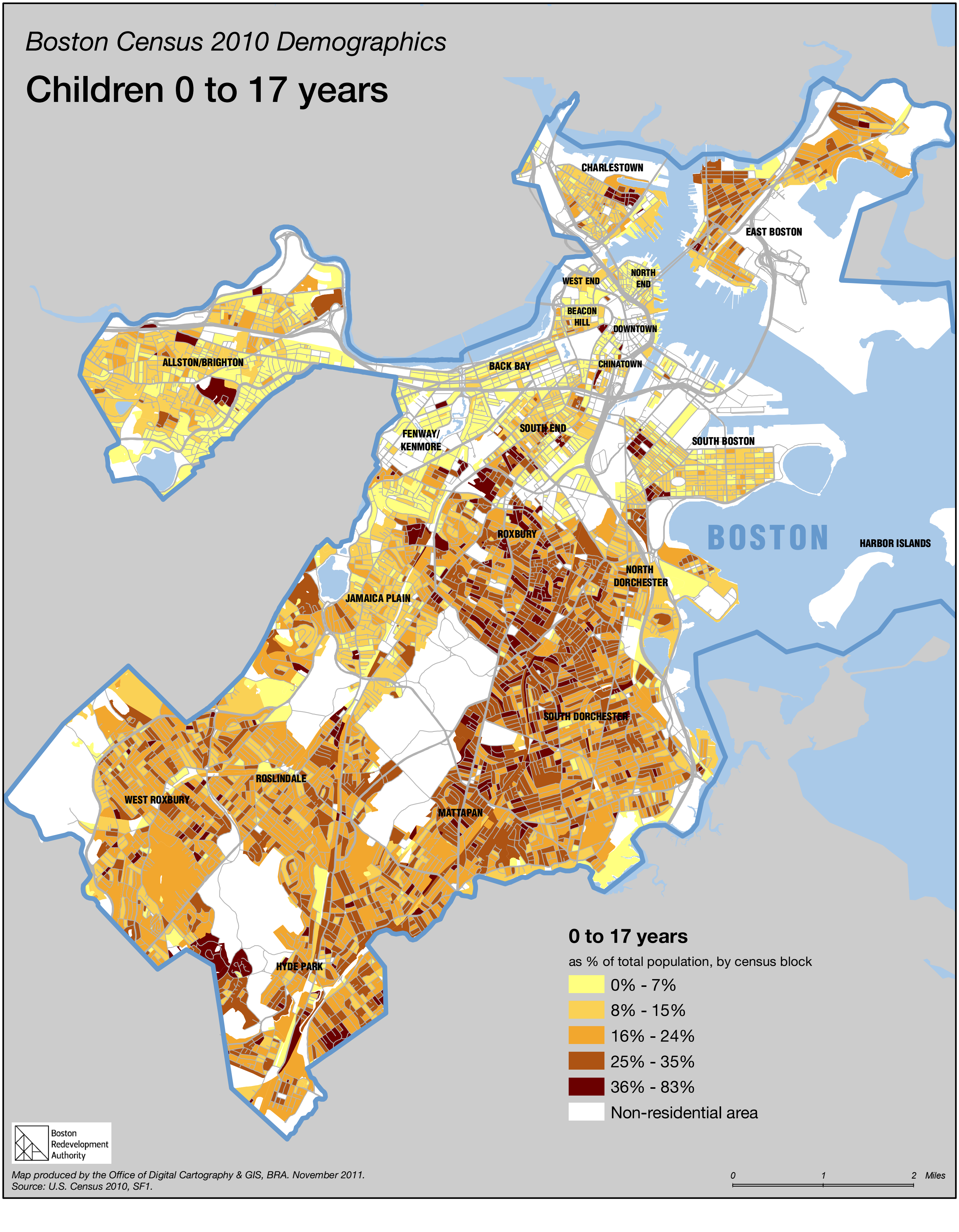

Figure A3. Comparison of ITT Estimates by Type of Crime, Boston versus Chicago

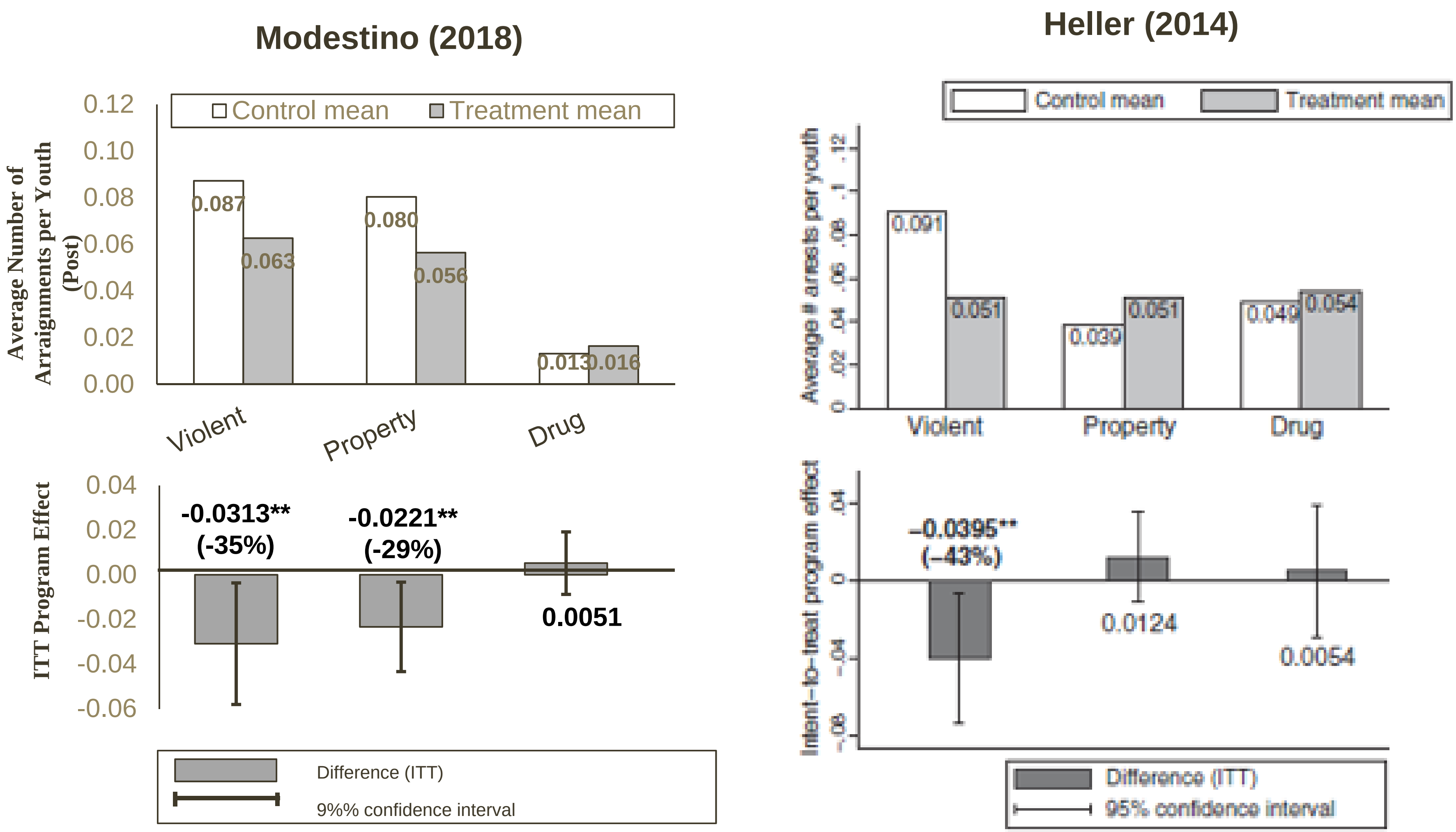

Figure A4. Comparison of Cumulative Reduction in Violent Crime, Boston versus Chicago

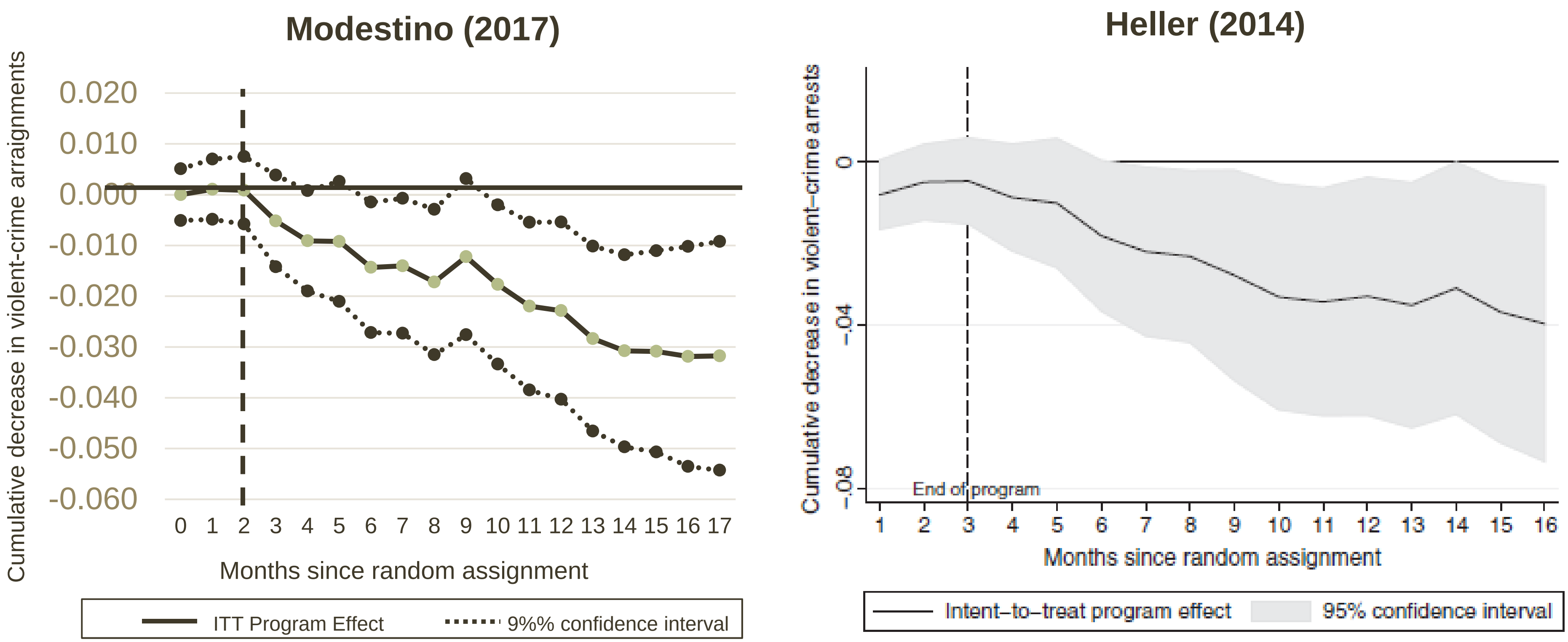

Table A1. ABCD Applicant Neighborhood by Lottery Outcome.

	Selected (Treatments)	Not Selected (Controls)		Census
Total selected by random assignment	1,186	3,049		
PERCENT IN EACH CATEGORY:				
Neighborhood				
Allston/Brighton	4.9%	5.1%		7.6%
Beacon Hill/Back Bay	0.4%	0.4%		2.0%
Charlestown	2.2%	2.4%		2.2%
Chinatown	0.8%	0.5%		1.7%
Dorchester	33.4%	32.8%		24.8%
East Boston	6.7%	6.6%		7.6%
Fenway	0.0%	0.0%		6.3%
Hyde Park	6.6%	6.5%		6.5%
Jamaica Plain	4.5%	4.7%		5.2%
Mattapan	9.1%	8.9%		4.7%
Mission Hill	1.8%	2.0%		2.6%
North End	0.1%	0.1%		0.4%
Roslindale	5.7%	5.7%		4.8%
Roxbury	10.3%	10.4%		11.4%
South Boston	6.4%	6.6%		3.7%
South End	6.1%	6.3%		3.1%
West End	0.1%	0.1%		0.4%
West Roxbury	0.8%	0.9%		5.1%

Source: Based on application data provided by the City of Boston Office of Workforce Development.

Table A2. ABCD Applicant Characteristics by Lottery Outcome versus Demographics from 5-Year ACS

	Selected (Treatments)	**Not Selected (Controls)**		**5-Year ACS**
Total selected by random assignment	1,186	3,049		
PERCENT IN EACH CATEGORY:				
Age				
14-17 years	79.4%	80.2%		28.3%
18-24 years	20.6%	19.8%		71.7%
Among those Age 14-17 years:				
Gender				
Female	53.1%	53.9%		51.5%
Male	46.9%	46.1%		48.5%
Race				
African American	51.3%	54.0%		50.1%
Asian	6.5%	5.0%		6.6%
White	9.6%	8.4%		9.5%
Other / Mixed-Race	32.5%	32.6%		33.8%

Source: SYEP based on application data provided by the City of Boston Office of Workforce Development.
5-Year ACS is from U.S. Census Bureau, 2011-2015 American Community Survey 5-Year Estimates.

Note: Sample for 5-Year ACS are low-income households with Income in the past 12 months below poverty level.

Table A3. Testing the Validity of the ABCD Lottery within Demographic Groups

	All groups combined	Youth: Age 14-18 years					
		African American		White		Other/two or more races	
		Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	0.012	-0.025	0.018	0.005	-0.059	-0.017	0.014
	(0.011)	(0.035)	(0.032)	(0.072)	(0.073)	(0.044)	(0.038)
Male	0.019	NA	NA	NA	NA	NA	NA
	(0.041)						
African American	-0.064	NA	NA	NA	NA	NA	NA
	(0.041)						
Asian	0.167 *	NA	NA	NA	NA	NA	NA
	(0.089)						
Other/two or more races	-0.001	NA	NA	NA	NA	NA	NA
	(0.044)						
English as preferred language	(0.048)	-0.130	0.412	-0.212	0.222	0.064	0.130
	(0.097)	(0.282)	(0.341)	(0.461)	(0.689)	(0.310)	(0.199)
Limited English ability	-0.003	-0.113	-0.269	0.461	-0.233	-0.097	-0.005
	(0.080)	(0.192)	(0.204)	(0.361)	(0.406)	(0.208)	(0.168)
In school	-0.043	0.114	0.147	0.083	-0.116	0.272	0.061
	(0.063)	(0.241)	(0.233)	(0.415)	(0.399)	(0.295)	(0.262)
Public assistance	0.063	0.004	-0.093	0.205	0.288	0.193	0.167
	(0.053)	(0.114)	(0.108)	(0.230)	(0.279)	(0.148)	(0.123)
Homeless	-0.018	-0.125	-0.308	0.199	0.028	-0.130	-0.082
	(0.082)	(0.216)	(0.198)	(0.290)	(0.388)	(0.264)	(0.210)
N	4235	891	1080	207	176	564	732

Source: Author's calculations based on application data provided by the City of Boston Office of Workforce Development.
Note: Robust standard errors are in parentheses. *Indicates difference is statistically significant at the 10 percent level.

Table A4. ABCD applicant characteristics by survey response.

	Treatment group All individuals		**Treatment group Responding to pre-/post-program survey**	
Number of youth	1,186		663	
Percent in each category:				
Age				
Mean	15.9	(0.058)	15.6	(0.083)
14-17 years	80.2%	(0.323)	80.7%	(0.013)
18-24 years	19.8%	(0.012)	19.3%	(0.013)
Gender				
Female	53.1%	(0.014)	54.7%	(0.016)
Male	46.9%	(0.014)	45.3%	(0.020)
Current education status				
In-school	87.6%	(0.010)	89.8%	(0.012)
Race				
African American	51.3%	(0.015)	54.6%	(0.016)
Asian	6.5%	(0.007)	6.5%	(0.010)
White	9.6%	(0.009)	9.9%	(0.012)
Other/Two or more races	32.5%	(0.014)	28.9%	(0.018)
Preferred language				
Chinese	0.2%	(0.001)	0.0%	(0.000)
English	95.1%	(0.006)	97.3%	(0.006)
Spanish	3.3%	(0.005)	1.7%	(0.005)
Other	1.4%	(0.003)	0.8%	(0.004)
Limited english ability				
Yes	7.1%	(0.007)	6.7%	(0.009)
Housing status				
Homeless	6.7%	(0.007)	6.4%	(0.006)
Household income type				
Public assistance	18.7%	(0.011)	17.6%	(0.015)
Disabled				
Yes	4.0%	(0.006)	3.7%	(0.037)

Source: Author's calculations based on survey data provided by the City of Boston, Office of Workforce Development.
Note: Standard errors are in parentheses. None of the differences are statistically significant.

Table A5. ABCD survey respondent characteristics by lottery outcome

	Treatment group responders		Control group responders		Difference
Total selected by random assignment	1,186		3,049		
Number of youth responding to survey	663		664		
Response rate	66.9%		21.8%		45.1
Age					
Mean	15.7	(0.078)	16.4	(0.081)	0.7 ***
14-17 years	80.2%	(0.014)	80.6%	(0.014)	-0.40
18-24 years	19.8%	(0.012)	19.4%	(0.012)	0.40
Gender					
Female	53.9%	(0.021)	65.2%	(0.021)	-11.38 ***
Male	46.1%	(0.021)	34.8%	(0.021)	11.38 ***
Race/ethnic group					
African American	51.5%	(0.021)	48.9%	(0.021)	2.63
Asian	6.5%	(0.010)	12.0%	(0.014)	-5.53 ***
White	3.2%	(0.007)	9.2%	(0.012)	-5.99 ***
Other/Two or more races	36.1%	(0.020)	26.8%	(0.019)	9.33 ***
Living situation					
Single parent family	63.7%	(0.020)	57.6%	(0.021)	6.17 **
Two parent family	29.4%	(0.019)	37.8%	(0.021)	-8.38 ***
Other relative	8.1%	(0.012)	10.7%	(0.013)	-2.62
Other	6.3%	(0.010)	4.4%	(0.009)	1.86
Language spoken at home					
Chinese	3.9%	(0.008)	5.5%	(0.010)	-1.59
English	74.0%	(0.019)	70.3%	(0.020)	3.67
Spanish	18.5%	(0.016)	10.7%	(0.013)	7.79 ***
Other	3.6%	(0.008)	13.5%	(0.015)	-9.88 ***

Source: Author's calculations based on survey data provided by the City of Boston, Office of Workforce Development.
Note: Standard errors are in parentheses. **Indicates difference is statistically significant at the 5 percent level, and*** at the 1 percent level.

Table A6. Item non-response rates for post-program survey: SYEP treatment group versus control group

CATEGORY	Treatment group (N=663)	Control group (N=664)	Difference (Percentage point)
Community engagement and social skills			
I have a lot to contribute to the groups I belong to	1.2%	0.9%	0.30
I feel connected to people in my neighborhood	1.2%	0.9%	0.30
I know how to manage my emotions and my temper	1.1%	1.5%	-0.45
I know how to ask for help when I need it	0.8%	0.6%	0.15
I know how to constructively resolve a conflict with a peer	0.6%	0.6%	0.00
I need to improve my conflict resolution skills	1.8%	1.8%	0.00
Job readiness skills			
I have all key information to apply for a job	2.3%	2.9%	-0.60
I have prepared a resume	2.6%	2.6%	0.00
I have prepared a cover letter	2.9%	3.0%	-0.15
I have developed answers to the usual interview questions	2.9%	2.7%	0.15
I have practiced my interviewing skills with an adult	3.9%	3.5%	0.46
I need to improve my job readiness skills	1.8%	1.8%	0.00
Future work plans and academic aspirations			
Plan to enroll in any eduation or training program after high school	2.4%	3.0%	-0.60

Source: Author's calculations based on survey data provided by the City of Boston Office of Workforce Development.

Note: None of the differences were statistically significant.

Table A7. Treatment-on-the-treated estimates from two-stage least squares regressions: Number of arraignments per youth

	First stage	Second stage		
	Depvar=PART	All crimes	Violent crimes	Property crimes
PART	-----	-0.041	-0.036 **	-0.029 *
	-----	(0.039)	(0.018)	(0.016)
SYEP	0.908 ***	-----	-----	-----
	(0.005)	-----	-----	-----
Baseline crime outcome (e.g. pre-arrest)	-0.025 **	1.844 ***	0.861 ***	0.782 ***
	(0.012)	-0.085	-0.054	-0.050
Age	-0.008 ***	0.014	-0.003	0.007
	(0.002)	(0.011)	(0.006)	(0.005)
Male	-0.011 **	0.152 ***	0.059 ***	0.039 **
	(0.005)	(0.033)	(0.017)	(0.014)
Black	-0.002	0.111 ***	0.055 ***	0.036 **
	(0.005)	(0.033)	(0.017)	(0.014)
Limited English	-0.016 *	0.028	-0.019	0.040
	(0.009)	(0.064)	(0.033)	(0.028)
In school	-0.017 *	0.119 *	-0.002	0.048
	(0.009)	(0.063)	(0.032)	(0.027)
Public assistance	0.007	0.015	0.014	0.010
	(0.006)	(0.044)	(0.023)	(0.019)
Homeless	-0.029 **	-0.006	0.004	0.035
	(0.013)	(0.096)	(0.050)	(0.042)
Disabled	-0.032 **	-0.065	-0.042	-0.007
	(0.013)	(0.090)	(0.046)	(0.039)
F-statistic	2915.260			
Number of Observations	4235	4235	4235	4235

Source: Author's calculations based on data provided by Massachusetts Department of Criminal Justice Information Services and Office of the Commissioner of Probation.
Note: *Indicates difference is statistically significant at the 10 percent level; ** at the 5 percent level; and*** at the 1 percent level.

Table A8. ITT program effect on criminal arraignments by type of crime: Percent of youth arraigned post-program

	Dependent Variable: Percent of youth arraigned			
	OLS Regressions		Probit Regressions	
	Without covariates	With covariates	Without covariates	With covariates
	(1)	(2)	(3)	(4)
All Crime	0.003	0.002	0.003	0.002
	(0.008)	(0.008)	(0.007)	(0.007)
Violent crimes	0.000	-0.001	0.000	-0.001
	(0.006)	(0.006)	(0.006)	(0.006)
Property crimes	-0.003	-0.004	-0.004	-0.003
	(0.005)	(0.005)	(0.005)	(0.005)
Drug crimes	0.000	0.000	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.002)
Gun crimes	-0.003	-0.003 *	-0.004	-0.005
	(0.002)	(0.002)	(0.003)	(0.004)
Other crimes	0.001	0.001	0.001	0.000
	(0.005)	(0.005)	(0.005)	(0.004)
Misdemeanors	0.002	0.001	0.002	0.001
	(0.007)	(0.007)	(0.006)	(0.006)
Felonies	-0.003	-0.004	-0.003	-0.003
	(0.006)	(0.006)	(0.007)	(0.006)
Includes baseline outcomes	No	Yes	No	Yes
Includes demographic characteristics	No	Yes	No	Yes
N	4235	4235	4235	4235

Source: Author's calculations based on data provided by the Massachusetts Department of Criminal Justice Information Services and Office of the Commissioner of Probation.

Note: Covariates include baseline outcomes and demographic characteristics such as age, gender, race/ethnicity, limited English, in school, public assistance, homelessness and disabled status. Coefficients for Probit regressions are marginal effects. Robust standard errors are in parentheses. *Indicates difference is statistically significant at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Table A9. ITT program effect on criminal arraignments by type of crime: Percent of youth arraigned post-program during 2016

	Dependent Variable: Percent of youth arraigned			
	OLS Regressions		Probit Regressions	
	Without covariates	With covariates	Without covariates	With covariates
	(1)	(2)	(3)	(4)
All Crime	0.005	0.005	0.006	0.005
	(0.006)	(0.006)	(0.006)	(0.005)
Violent crimes	0.000	0.000	0.001	0.000
	(0.036)	(0.005)	(0.004)	(0.004)
Property crimes	-0.001	-0.001	-0.001	0.000
	(0.004)	(0.004)	(0.004)	(0.004)
Drug crimes	0.001	0.000	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Gun crimes	-0.002	-0.002	-0.002	-0.003
	(0.002)	(0.002)	(0.003)	(0.003)
Other crimes	0.004	0.004	0.003	0.003
	(0.004)	(0.004)	(0.003)	(0.003)
Misdemeanors	0.004	0.003	0.003	0.003
	(0.005)	(0.005)	(0.005)	(0.005)
Felonies	-0.002	-0.002	0.000	0.000
	(0.005)	(0.005)	(0.005)	(0.005)
Includes baseline outcomes	No	Yes	No	Yes
Includes demographic characteristics	No	Yes	No	Yes
N	4235	4235	4235	4235

Source: Author's calculations based on data provided by the Massachusetts Department of Criminal Justice Information Services and Office of the Commissioner of Probation.
Note: Covariates include baseline outcomes and demographic characteristics such as age, gender, race/ethnicity, limited English, in school, public assistance, homelessness and disabled status. Coefficients for Probit regressions are marginal effects. Robust standard errors are in parentheses. *Indicates difference is statistically significant at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Table A10. Pre vs. post Program comparisons for criminal justice outcome measures: Treatments vs. controls.

	Treatment group					
	Number of arraignments per youth			**Percent of youth arraigned for a criminal charge**		
	Pre	Post	Diff: Post-Pre	Pre	Post	Diff: Post-Pre
All crimes	0.12	0.18	0.06 **	4.1%	5.1%	1.1 **
Violent crimes	0.07	0.06	-0.01	3.2%	3.9%	0.7 *
Property crimes	0.04	0.06	0.01	3.1%	3.7%	0.6 *
Drug crimes	0.00	0.02	0.02	0.7%	0.9%	0.2
Gun crimes	0.00	0.02	0.02	0.4%	0.8%	0.4
Other crimes	0.01	0.03	0.02	1.6%	2.2%	0.5 *
Misdemeanor	0.06	0.09	0.03 *	3.6%	4.2%	0.6 *
Felony	0.05	0.09	0.03 *	3.6%	4.4%	0.7 *
	Control group					
	Number of arraignments per youth			**Percent of youth arraigned for a criminal charge**		
	Pre	Post	Diff: Post-Pre	Pre	Post	Diff: Post-Pre
All crimes	0.14	0.21	0.08 **	3.6%	5.4%	1.8 **
Violent crimes	0.06	0.09	0.03 **	2.8%	4.3%	1.4 **
Property crimes	0.05	0.08	0.03 **	2.1%	3.0%	0.9 **
Drug crimes	0.01	0.01	0.00	0.4%	0.6%	0.2
Gun crimes	0.00	0.01	0.01	0.2%	0.4%	0.2
Other crimes	0.02	0.02	0.00	1.2%	1.9%	0.7
Misdemeanor	0.08	0.11	0.04 **	3.2%	4.4%	1.2 **
Felony	0.05	0.10	0.04 **	2.9%	4.4%	1.4 **

Source: Author's calculations based on administrative records from the Massachusetts Department of Criminal Justice Information Services and Office of the Commissioner of Probation.

Note: Pre-program is defined as the 17 months prior to random assignment (February 2014 through June 2015) and post-program is defined as the 17 months afer the program ends (September 2015 though November 2016).

*Indicates difference is statistically significant at the 10 percent level; ** at the 5 percent level; and*** at the 1 percent level.

Table A11. ITT program effect on number of criminal arraignments: Number of post-program arraignments for youth arraigned during the post-period

	Dependent Variable: Number of arraignments per youth				
	Unadjusted Means			Regression Adjusted	
	Treatment	Control	Difference	OLS	Poisson
All crimes	3.070	4.380	-1.310 **	-1.171 **	-1.300 **
	(0.397)	(0.335)	(0.624)	(0.526)	(0.612)
Violent crimes	1.140	1.891	-0.752 **	-0.798 **	-0.859 **
	(0.175)	(0.202)	(0.365)	(0.279)	(0.351)
Property crimes	0.674	1.302	-0.628 *	-0.539 **	-0.643 **
	(0.144)	(0.193)	(0.346)	(0.268)	(0.251)
Drug crimes	0.326	0.240	0.085	0.194	0.130
	(0.148)	(0.067)	(0.144)	(0.150)	(0.148)
Gun crimes	0.349	0.349	0.000	0.038	-0.003
	(0.305)	(0.090)	(0.234)	(0.303)	(0.315)
Other crimes	0.581	0.597	-0.016	0.047	0.044
	(0.167)	(0.086)	(0.178)	(0.200)	(0.195)
Misdemeanor	1.628	2.372	-0.744 *	-0.672 *	-0.733 *
	(0.258)	(0.229)	(0.424)	(0.363)	(0.411)
Felony	1.442	1.992	-0.550 *	-0.471	-0.564
	(0.254)	(0.172)	(0.333)	(0.297)	(0.362)
Includes baseline outcomes	----	----	----	Yes	Yes
Includes demographic characteristics	----	----	----	Yes	Yes
N	43	129	172	172	172

Source: Author's calculations based on administrative records from the Massachusetts Department of Criminal Justice Information Services and Office of the Commissioner of Probation.

Note: Covariates include baseline outcomes and demographic characteristics such as age, gender, race/ethnicity, limited English, in school, public assistance, homelessness and disabled status. Coefficients for Poisson regressions are marginal effects. Standard errors are in parentheses with robust standard errors for regression estimates. *Indicates difference is statistically significant at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Table A12. ITT program effect on number of criminal arraignments: Number of post-program arraignments for youth ever arraigned prior to the program versus youth with no prior arraignments.

	Youth EVER arraigned prior to the program				Youth NEVER arraigned prior to the program	
	Re-arraignment Rate		**Number of Post-Program Arraignments**		**Number of Post-Program Arraignments**	
	Treatment	**Control**	OLS	Poisson	OLS	Poisson
All crimes	0.344	0.351	-0.361	-0.443	-0.012	-0.013
	(0.061)	(0.033)	(0.422)	(0.481)	(0.020)	(0.022)
Violent crimes	0.180	0.227	-0.335 **	-0.441 **	-0.003	-0.003
	(0.050)	(0.029)	(0.158)	(0.212)	(0.012)	(0.012)
Property crimes	0.115	0.161	-0.263 *	-0.361 *	-0.004	-0.005
	(0.041)	(0.025)	(0.159)	(0.217)	(0.007)	(0.008)
Drug crimes	0.082	0.062	0.130	0.139	-0.001	-0.013
	(0.035)	(0.017)	(0.112)	(0.092)	(0.001)	(0.010)
Gun crimes	0.033	0.052	0.105	0.077	-0.005 **	-0.006 **
	(0.023)	(0.015)	(0.209)	(0.147)	(0.002)	(0.003)
Other crimes	0.164	0.147	0.059	0.048	0.000	0.000
	(0.048)	(0.024)	(0.135)	(0.127)	(0.005)	(0.005)
Misdemeanor	0.295	0.280	-0.200	-0.228	-0.010	-0.011
	(0.059)	(0.031)	(0.260)	(0.298)	(0.013)	(0.015)
Felony	0.230	0.284	-0.098	-0.143	-0.002	-0.002
	(0.054)	(0.031)	(0.216)	(0.246)	(0.010)	(0.010)
Includes baseline outcomes	-----	-----	Yes	Yes	Yes	Yes
Includes demographic characteristics	-----		Yes	Yes	Yes	Yes
N	61	211	272	272	3963	3963

Source: Author's calculations based on administrative records from the Massachusetts Department of Criminal Justice Information Services and Office of the Commissioner of Probation.

Note: Covariates include baseline outcomes and demographic characteristics such as age, gender, race/ethnicity, limited English, in school, public assistance, homelessness and disabled status. Coefficients for Poisson regressions are marginal effects. Standard errors are in parentheses with robust standard errors for regression estimates. *Indicates difference is statistically significant at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Table A13. Relationship between SYEP impact on changes in community engagement and social skills and subsequent criminal activity: Number of arraignments per youth

	Any Improvement				**Meaningful Improvement**			
Dummy variable indicating improvement in:	Violent crimes		Property crimes		Violent crimes		Property crimes	
Community engagement and social skills	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
<u>OLS Regressions</u>								
Contributing to the groups they belong to	-0.012	(0.011)	-0.004	(0.011)	-0.028	(0.015) *	-0.031	(0.008) ***
Connecting to people in their neighborhood	-0.001	(0.012)	0.008	(0.012)	-0.031	(0.015) **	-0.031	(0.008) ***
Managing emotions	-0.031	(0.011) ***	-0.021	(0.011) **	-0.051	(0.011) ***	-0.030	(0.008) ***
Asking for help	0.004	(0.011)	-0.017	(0.011)	-0.029	(0.015) *	-0.025	(0.012) **
Resolving conflict with a peer	-0.048	(0.023) ***	-0.025	(0.010) **	-0.051	(0.011) ***	-0.030	(0.008) ***
<u>Poisson Regressions</u>								
Contributing to the groups they belong to	0.000	(0.021)	-0.009	(0.026)	-0.032	(0.019) *	-0.037	(0.017) **
Connecting to people in their neighborhood	0.011	(0.028)	0.014	(0.030)	-0.036	(0.020) *	-0.037	(0.017) **
Managing emotions	-0.044	(0.021) ***	-0.038	(0.018) **	-0.066	(0.019) ***	-0.040	(0.018) **
Asking for help	-0.017	(0.020)	-0.020	(0.018)	-0.031	(0.019) *	-0.028	(0.017) *
Resolving conflict with a peer	-0.057	(0.027) ***	-0.041	(0.021) **	-0.060	(0.020) ***	-0.046	(0.016) **
Number of Observations	4235		4235		4235		4235	

Source: Author's calculations based on data provided by Massachusetts Department of Criminal Justice Information Services and Office of the Commissioner of Probation.
Note: Covariates include baseline outcomes and demographic characteristics such as age, gender, race/ethnicity, limited English, in school, public assistance, homelessness and disabled status. Robust standard errors are in parentheses. See Table A5 in the appendix for the baseline treatment-on-the treated results. *Indicates difference is statistically significant at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Table A14. Relationship between short-term behavioral changes and subsequent criminal activity: SYEP participants

Dummy variable indicating improvement in:	**Number of arraignments per youth**			
	Violent cimes		Property crimes	
	Coefficient	Standard error	Coefficient	Standard error
Community engagement and social skills				
Contributing to the groups they belong to	-0.007	(0.012)	-0.003	(0.011)
Connecting to people in their neighborhood	0.007	(0.004)	0.019	(0.012)
Managing emotions	-0.031	(0.013) **	-0.021	(0.011) *
Asking for help	0.005	(0.012)	-0.025	(0.011) **
Resolving conflict with a peer	-0.049	(0.023) **	-0.042	(0.021) **
Improving conflict resolution skills (overall)	-0.109	(0.043) **	-0.053	(0.031) *
Job readiness skills				
Having key information to apply for a job	0.001	(0.019)	0.013	(0.015)
Preparing a resume	0.021	(0.013)	-0.015	(0.011)
Preparing a cover letter	0.003	(0.013)	-0.010	(0.010)
Developing answers to interview questions	-0.033	(0.019) *	-0.015	(0.016)
Practicing interviewing with an adult	0.019	(0.012)	0.020	(0.012)
Improving job readiness skills (overall)	-0.017	(0.013)	-0.014	(0.012)
Academic aspirations				
Planning to attend a four-year college	-0.005	(0.014)	0.002	(0.013)
Includes baseline outcomes	Yes		Yes	
Includes demographic characteristics	Yes		Yes	
N	663		663	

Source: Author's calculations based on data provided by Massachusetts Department of Criminal Justice Information Services and Office of the Commissioner of Probation.

Note: Sample includes participants that answered both the pre- and post-survey. Covariates include baseline outcomes and demographic characteristics such as age, gender, race/ethnicity, limited English, in school, public assistance, homelessness and disabled status. Robust standard errors are in parentheses. *Indicates difference is statistically significant at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Table A15. Pre vs. post program comparisons for criminal justice outcome measures: Youth Options Unlimited

	Treatment group						
	Number of arraignments per youth			Percent of youth arraigned for a criminal charge			Re-arraignment rate
	Pre	Post	Diff: Post-Pre	Pre	Post	Diff: Post-Pre	
Percent of youth arraigned for a criminal charge							
All crimes	4.70	1.20	-3.50 **	41.6%	21.2%	-20.4 **	44.7%
Violent crimes	1.87	0.38	-1.49 *	31.9%	18.6%	-13.3 **	52.8%
Property crimes	1.19	0.34	-0.86	32.7%	19.5%	-13.3 **	54.1%
Drug crimes	0.32	0.07	-0.25	10.6%	7.1%	-3.5	66.7%
Gun crimes	0.58	0.19	-0.38	11.5%	7.1%	-4.4	61.5%
Other crimes	0.69	0.19	-0.50	23.0%	15.9%	-7.1	61.5%
Misdemeanor	2.39	0.53	-1.86 **	38.9%	21.2%	-17.7 **	47.7%
Felony	2.14	0.50	-1.64 *	38.1%	20.4%	-17.7 **	48.8%
Number of youth	119	119	0	119	119	0	119

Source: Author's calculations based on administrative records from the Massachusetts Department of Criminal Justice Information Services and Office of the Commissioner of Probation.

Note: P re-program is defined as the 17 months prior to random assignment (February 2014 through June 2015) and post-program is defined as the 17 months afer the program ends (September 2015 though November 2016). To determine if an individual has ever been charged with a crime, the pre-program period is defined as any time prior to July 2015 and the post-program period is defined as September 2015 through November 2016. The re-arraignment rate is calculated as the number of youth re-arraigned during the post-program period divided by the number of youth arraigned during the pre-program period. *Indicates difference is statistically significant at the 10 percent level; ** at the 5 percent level; and*** at the 1 percent level.

Table A16. Estimated reduction in costs associated with violent and property crimes

	Second Stage		
	Violent crimes	Violent crimes trimmed	Property crimes
PARTICIPANT	-4733.25	-1793.38 *	-135.36 **
	(3704.20)	(1061.77)	(68.28)
R-squared	0.002	0.021	0.024
N	4235	4235	4235

Source: Author's calculations based on data provided by Massachusetts Department of Criminal Justice Information Services and Office of the Commissioner of Probation.

Note: Cost estimates based on treatment-on-the-treated estimates from two-stage least squares regressions. *Indicates difference is statistically significant at the 10 percent level and ** at the 5 percent level.