

Rethinking Youth Employment Programs: Evidence from Two Summer Jobs Experiments

Sara B. Heller
University of Pennsylvania

Please note:

The talk will be based on the attached published paper and related subsequent work. Unfortunately, the subsequent work is not currently legally authorized to circulate in written form, although I am cleared to present it.

This is the author's version of the work prior to copyediting and layout. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in Science Magazine, Vol 346, 5 December 2014. doi: [10.1126/science.1257809](https://doi.org/10.1126/science.1257809)

Summer Jobs Reduce Violence among Disadvantaged Youth

Authors: Sara B. Heller^{1*}

Affiliations:

¹ University of Pennsylvania and University of Chicago Crime Lab.

*Correspondence to: Sara Heller, hellersa@sas.upenn.edu.

Abstract:

Every day, acts of violence injure over 6,000 people in the U.S. Despite decades of social science arguing that joblessness among disadvantaged youth is a key cause of violent offending, programs to remedy youth unemployment do not consistently reduce delinquency. This study tests whether summer jobs, which shift focus from remediation to prevention, can reduce crime. In a randomized controlled trial among 1,634 disadvantaged high school youth in Chicago, assignment to a summer jobs program decreases violence by 43 percent over 16 months (3.95 fewer violent-crime arrests per 100 youth). The decline occurs largely after the 8-week intervention ends. The results suggest the promise of using low-cost, well-targeted programs to generate meaningful behavioral change, even with a problem as complex as youth violence.

Main Text:

Every day in the U.S., acts of violence kill almost 150 people and injure over 6,000 more (1). (2) This "public health crisis" (3) disproportionately involves youth, who are twice as likely as adults to be both victims and perpetrators of violence. (4) The problem is most concentrated among disadvantaged minority youth; violent-crime arrest rates for African-American juveniles are five times as high as their white counterparts (5).

Poverty scholar William Julius Wilson identifies one major cause of these racial violence disparities among young people: joblessness (6). His argument adds to decades of social science investigating how poor job prospects cause crime, from weakening social bonds to generating psychic strain to reducing the perceived cost of punishment (7-10). Policy discussions often conclude that public employment and training programs are a solution to youth violence on the grounds that "nothing stops a bullet like a job" (11).

The empirical literature on youth employment programs, however, suggests the practical applications of this idea are limited. (12) Several major employment programs fail to reduce delinquency among youth (13, 14). The two experimentally-evaluated interventions that do lower crime, at least during the program, involve such intensity and expense that their benefits fail to outweigh their costs (15, 16). Attempts to provide shorter, more scalable employment services have, if anything, increased adolescents' criminal behavior (17). The theoretical promise of using employment interventions to fight youth violence does not appear to translate easily into empirical reality; only very high levels of investment seem able to reduce crime (supplementary materials, section 2.1). (18)

But one widespread type of youth employment program has not yet been rigorously evaluated: summer jobs (19, 20). Despite some promising observational studies (21), there is little convincing causal evidence on the effects of these short-term, low-cost programs and, to my knowledge, no experimental evidence on whether summer jobs can reduce crime (supplementary materials, section 2.1).

This paper presents the results of a large-scale randomized controlled trial testing a public summer jobs program in Chicago called One Summer Plus (OSP). OSP offers 8 weeks of part-time summer employment at Illinois' minimum wage (\$8.25/hour). Local community organizations place youth in non-profit and government jobs (e.g., summer camp counselors, workers in a community garden, office assistants for an alderman). Youth are assigned job mentors – adults who help them learn to be successful employees and to navigate barriers to employment – at a ratio of about 10 to 1. Half the treatment group also receives social-emotional learning (SEL) based on cognitive behavioral therapy principles, aimed at teaching youth to understand and manage the aspects of their emotions and behavior that might interfere with employment (supplementary materials, section 2.2).

The study randomly assigns 1,634 8th – 12th grade applicants enrolled in 13 high-violence Chicago schools to program (jobs only or jobs + SEL) or control conditions. (22) Youth in the jobs-only group are offered 25 hours per week of paid employment; youth in the jobs + SEL group are paid for 15 hours of work and 10 hours of SEL weekly. Control youth are excluded from the program but free to pursue other opportunities.

The existing employment literature might suggest that this kind of low-dosage jobs program is unlikely to change criminal behavior. But there is an important difference between summer jobs

for in-school youth and the previous literature: The well-studied youth employment programs generally act as tertiary prevention, targeting youth already out of school and struggling in the labor market. Research in domains like education and health, however, suggests that for many negative outcomes caused at least in part by prior behavior (e.g., poor school performance and dropout, certain types of cancer), primary and secondary prevention – intervening before onset rather than managing or trying to reverse a problem once it occurs – can improve outcomes more effectively with less intensive treatment (23, 24). (25)

If this well-established idea that prevention can be more effective than remediation also applies to the employment domain, teaching adolescents how to be successful employees, facilitating connections to employer networks, and providing work experience before they drop out of school might reduce crime with less intensive intervention than is required for already disconnected youth. Offering summer employment at this key point in the life course could make crime a relatively less attractive option, strengthen social bonds, and develop “soft” skills like self-efficacy and impulse control (7, 26-28). Waiting until after dropout to intervene may make it more difficult to reduce violence in particular, given the apparently causal relationship between school dropout and criminal behavior, especially murder and assault (29). Limiting the intervention to summer avoids direct conflicts between work and school. Summer jobs also provide wages and structured activity during a high-crime season when youth might otherwise be idle, both of which may affect crime (30, 31).

Note, however, that the theory on summer jobs is not entirely clear-cut. Additional income could be spent on crime-inducing goods like drugs and alcohol, and time spent traveling to and from work might increase exposure to criminal opportunities. Even the attempt to keep youth busy (the “incapacitation effect”) may be poorly targeted, since most jobs are during business hours, while most crime occurs during evenings and weekends. More broadly, a short-term subsidized job might not be enough to generate behavior changes beyond the summer itself. The effects of summer jobs are therefore an empirical question.

Data

Data come from matching study youth to administrative data sources (supplementary materials, materials and methods). Program participation is from provider-tracked attendance records. Student-level administrative records from the Chicago Public Schools capture pre- and post-program academic outcomes from the 2011-12 and 2012-13 academic years, respectively. Demographic information on applicants’ neighborhoods comes from matching the Census tract of youths’ home addresses to the 2010 American Community Survey. The main outcome measures are from individual-level Chicago Police Department arrest records covering both juveniles and adults. (32) The study uses data through September 25, 2013, which is 16 months after randomization (13 months after the end of the program).

The analysis categorizes arrests by offense type (violent, property, drug, and other). It is common for social interventions to have differential impacts on these different types of crime (31, 33, 34), likely because some of the underlying causes differ. Violent crime stems from conflicts between people, so problematic cognitive and emotional responses to social interactions – including hostile attribution bias, uncontrolled anger, and “hot” decision-making – are thought to be proximal causes of youth violence in particular (35, 36). Non-violent crimes, which involve property or drugs more often than interpersonal conflict, may be relatively more responsive to situational and economic factors (37, 38). Since OSP could affect how youth perceive and

respond to social interactions differently from how it affects their economic situations, opportunities for crime, or drug use, the study estimates program effects separately by crime type.

Descriptive Statistics

Table 1 shows average applicant characteristics on a variety of pre-program measures. None of the differences across the columns is statistically significant, nor are they jointly significant ($F(31, 1517) = 0.37, p = 0.9995$), confirming the success of randomization.

Applicants are on average just over age 16 (minimum = 14, maximum = 21), and almost all are African-American. Over 90 percent are eligible for free or reduced price lunches (a proxy for family poverty), with average GPAs around a C in the prior fall semester. Study youth missed 18 percent of the pre-program school year, or around 6 weeks. About 20 percent had been arrested at baseline, and just over 20 percent had been victimized. Applicants live in highly disadvantaged neighborhoods: unemployment averages over 19 percent; a third of households are under the poverty line; and violent crime rates are extremely high (over 2,100 incidents per 100,000 people).

Results

The main text presents intent-to-treat (ITT) estimates, which measure the average difference between those randomly assigned to treatment and control groups. Three-quarters of treatment youth actually participate (supplementary materials, section 2.3, Table S1). Because not all youth offered the program enroll, the ITT understates the effect of the program on those who choose to participate. This treatment-on-the-treated effect is discussed and estimated in the methods section and section 2.4 of the supplementary materials. Program impacts are shown both for the treatment group as a whole and each treatment arm.

During the 16-month follow-up period, about 17 percent ($n=155$) of the control group is arrested for any crime, with an average of 0.30 arrests per youth. Figure 1 presents the main ITT results. Violent-crime arrests among the treatment group decrease by 43 percent relative to the control group (0.0395 fewer arrests, or almost 4 fewer per 100 youth, $p = 0.022$). There are no significant changes in other types of arrests. (39) The results are robust to accounting for the number of hypothesis tests conducted, allowing youth outcomes to be correlated within schools, and using a non-linear specification for count data (supplementary materials, section 2.5). Although the study was not powered to detect heterogeneous treatment effects across subgroups, there is suggestive evidence that arrests fall more among youth at higher risk of violence (supplementary materials, section 2.7).

An important question concerns what mechanisms drive this considerable behavioral change. One hypothesis for why prior youth employment programs require high intensity to succeed is that disadvantaged adolescents may lack the “soft skills” to engage with less intensive pro-social programming (40). If so, OSP’s SEL curriculum, which focuses on emotional and conflict management, social information processing, and goal setting, could be a key driver of the violence decrease. Because SEL was randomly assigned, the study can separately identify any additional role of SEL.

Table 2 shows ITT violence effects by treatment group. Not only is the difference between groups statistically insignificant, but the magnitude is also tiny (1 percent of the control mean). The nearly identical point estimates suggest that the statistical similarity across groups is not just

the result of limited power to detect subgroup differences. Instead, it appears that both groups of youth experienced very similar drops in violence. The same is true for other crime types (supplementary materials, section 2.6, Table S4).

Although the experimental design can only directly isolate the role of SEL, another potentially important mechanism is testable indirectly: whether the decrease in violence is a mechanical result of youth having less time to engage in crime while working over the summer (an “incapacitation” effect). If so, and youth returned to their prior behavioral patterns immediately after the program ended, one would expect to see a large violence drop during the program period, after which there would be no treatment-control difference. This is not the case. The ITT violence decrease during the 3 months between random assignment and the end of the program (not shown) is proportionally large but statistically insignificant (-0.005 violent-crime arrests, $p = 0.39$, control mean = 0.01). During the following 13 months (excluding the program period), treatment reduces the number of violent-crime arrests by -0.035 ($p = 0.03$, control mean = 0.08).

Figure 2 shows the time path of the violence decrease in greater detail. It graphs the cumulative treatment effect over time, with each point adding an additional month of data to the prior effect. The drop in arrests becomes statistically different from zero around month 6, three months after the end of the program, and continues to grow through month 11, after which it flattens out. The downward slope of the effect makes it clear that the bulk of the drop in violence accrues between months 5 and 11, well after the end of the program at month 3.

One way the program might have changed behavior during these later months is by increasing time spent in school, either by developing pro-social attitudes or providing information on how the labor market values education. However, schooling data suggest this is unlikely to be the key mechanism. There is no treatment effect on days present (or other academic outcomes) during the following school year, with confidence intervals small enough to make changes in attendance an implausible cause of the drop in violence (supplementary materials, section 2.10, Table S7).

Discussion

Prior research on youth employment suggests that only costly and intensive employment programs can reduce crime. The current study demonstrates that when offered to youth still in school, an intervention need not be lengthy to change behavior; an 8-week summer jobs program reduces violence among an adolescent population living in some of the most violent neighborhoods in the country.

The results are consistent with the idea that employment programs can have a larger impact at lower cost when offered to disadvantaged youth prior to school exit. There are, however, other possible reasons why this study finds larger effects than previous work. For example, OSP involved an adult job mentor making regular visits to the workplace, while most of the evaluated U.S. youth programs do not. The measurement of crime may also matter. If reaching youth prior to school exit is the key difference, one might expect multi-year interventions for high school youth that aim to improve both education and employment outcomes to generate even larger impacts than seen here. In fact, they generally find no effects on crime (41, 42). But the evaluations of those programs use self-reported overall crime rates, separating only drug and alcohol use. Given that effects are heterogeneous by crime type, it is possible that these programs also reduced violence, but that the studies aggregated crime outcomes to a level that masked the effect.

The estimates in this paper rely on administrative arrest records, which only capture offenses known to the police. Nationally, about half of violent incidents are reported to police, about half of which result in an arrest (43, 44). As such, the estimates here may considerably understate the number of violent crimes prevented (supplementary materials, section 2.9).

The fact that both the jobs-only and jobs + SEL versions of the program were equally effective means that SEL cannot be the only operative mechanism. A similar SEL program has reduced violence on its own (45), but the jobs-related programming was enough to generate the steep drop in violence independently. Because replacing 2 hours of work with an SEL curriculum neither improved nor reduced the intervention's impact, the relevant mechanism is likely to be something common across the two types of programming.

One possibility is that the substance of the SEL curriculum – teaching youth to process social information, manage emotions, and set and achieve goals more successfully – was taught equally well on-the-job. Youth could have learned these skills from interactions with the job mentors, who were part of both treatment arms. Reports from the program providers suggest that the mentors play a large part in helping youth learn to manage conflict in the workplace, in addition to teaching basic job skills. The process of working itself might also be enough to improve self-control, develop self-efficacy, and reduce frustration (all potential determinants of how youth perceive and respond to conflict). One study, which randomly assigned adults to an earnings subsidy program to increase employment, found that treated individuals report a greater sense of control over their lives and less anger about their lack of opportunities (46).

Another possibility is that the relevant mechanism was something else that both groups received: income, a caring adult, job skills, and/or employer connections. The experimental design cannot tease out which program element or elements generate the decrease in violence. Nonetheless, since schooling outcomes show little movement, it is unlikely that changes in perceptions about the value of schooling, or simply the amount of time in school, are the main drivers of the violence drop. The program provides a relatively large income shock, averaging \$1,400 in neighborhoods where one-third of households are below the poverty line and median income is about \$35,000. Additional income could change criminal behavior directly or increase parental supervision by reducing how much parents need to work away from home. It is also possible that improved post-program employment plays a role, though the direction of that effect is hard to predict. Working more could reduce all types of crime through incapacitation, or by making it more costly to be incarcerated. But the increased income could also have mixed effects, such as reducing economically motivated crimes like theft while increasing the ability to purchase drugs.

The fact that OSP only reduces violent crime is perhaps most consistent with a role for improvements in self-control, social information processing, and decision-making, which as discussed above, are more central to violent behavior than other types of crime. Other interventions that target these skills through either explicit curricula or mentorship have also reduced youth delinquency with very few contact hours (47, 48).

One key insight from the data is that incapacitation alone – the mechanical reduction in free time during the summer – cannot explain the program's success. Violence reductions are large and statistically significant during the 13 months after the program is over, and the point estimate is 7 times larger at the end of the follow-up period than at the end of the summer. This pattern is not consistent with the idea that behavior only changed during the initial 8 weeks of the program; something about youths' summer experiences also changed their future behavior.

As in any social experiment, external validity is not guaranteed. The study sample is fairly representative of low-income, African-American youth living in urban neighborhoods who are still enrolled in school. Many other cities have successfully implemented summer jobs programs, suggesting that the basic approach is one that may not be difficult to scale (although displacement effects which reduce the number of jobs available to non-program youth become a bigger concern in larger-scale programs). A better understanding of how broadly these results will generalize requires replications in other cities with different service environments and youth populations. Policymakers should proceed with caution until future studies can establish exactly what works for whom, how long effects persist, and whether the program's benefits outweigh its costs (supplementary materials, section 2.11).

In the meantime, this study provides causal evidence on the effects of a widespread but understudied intervention. The results echo a common conclusion in education and health research: that public programs might do more with less by shifting from remediation to prevention. The findings make clear that such programs need not be hugely costly to improve outcomes for disadvantaged youth; well-targeted, low-cost employment policies can make a substantial difference, even for a problem as destructive and complex as youth violence.

References and Notes:

1. Center for Disease Control and Prevention. (Web-based Injury Statistics Query and Reporting System (WISQARS), 2014).
2. Average of 5 most recent years of violence-related fatalities (from death certificates) and non-fatal injuries (from emergency room reports)
3. Center for Disease Control and Prevention. (STRYVE to Prevent Youth Violence Website, <http://www.cdc.gov/features/youthviolence/index.html>, 2014).
4. Authors' calculations for ages 10-24 versus over 24 from 2012 FBI Uniform Crime Reports, Census data, and (1)
5. Office of Juvenile Justice and Delinquency Prevention. (OJJDP Statistical Briefing Book, U.S. Department of Justice, Washington DC, 2014, http://www.ojjdp.gov/ojstatbb/crime/JAR_Display.asp?ID=qa05261).
6. W. J. Wilson, *When work disappears: The world of the urban poor*. (Alfred Knopf, New York, 1996).
7. G. S. Becker, Crime and Punishment: An Economic Approach. *J. Polit. Economy* **76**, 169-217 (1968).
8. T. P. Thornberry, R. L. Christenson, Unemployment and criminal involvement: An investigation of reciprocal causal structures. *Am. Sociol. Rev.* **49**, 398-411 (1984).
9. R. J. Sampson, J. H. Laub, Crime and deviance over the life course: The salience of adult social bonds. *Am. Sociol. Rev.* **55**, 609-627 (1990).
10. R. K. Merton, Social structure and anomie. *Am. Sociol. Rev.* **3**, 672-682 (1938).
11. P. J. Cook, J. Ludwig, in *Controlling Crime: Strategies and Tradeoffs*, P. J. Cook, J. Ludwig, J. McCrary, Eds. (University of Chicago Press, Chicago, 2011), pp. 1-39.
12. The focus here is on youth programs, because disadvantaged youth are at such disproportionate risk of violent offending. There have been many public efforts to provide vocational or on-the-job training, job search assistance, and remedial education to adults as well. Those programs' crime effects are also mixed (49), perhaps in part because the largest employment impacts tend to be among adult women – a group with very low levels of criminal involvement (supplementary materials, section 2.1).
13. M. Millenky, D. Bloom, S. Muller-Ravett, J. Broadus, "Staying on Course: Three-Year Results of the National Guard Youth Challenge Evaluation" (MDRC, New York, 2011).
14. C. Uggen, Work as a turning point in the life course of criminals: A duration model of age, employment, and recidivism. *Am. Sociol. Rev.* **65**, 529-546 (2000).
15. P. Z. Schochet, J. Burghardt, S. McConnell, Does Job Corps work? Impact findings from the National Job Corps Study. *Am. Econ. Rev.* **98**, 1864-1886 (2008).
16. G. Cave, H. Bos, F. Doolittle, C. Toussaint, "JOBSTART. Final Report on a Program for School Dropouts," (Manpower Demonstration Research Corporation (MDRC), New York, 1993).
17. H. S. Bloom, L. L. Orr, S. H. Bell, G. Cave, F. Doolittle, W. Lin, J. M. Bos, The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act study. *J. Hum. Resour.* **32**, 549-576 (1997).
18. Although there is a consensus that generating measurable changes in youth outcomes requires comprehensive and expensive intervention, conclusions about how much investment is desirable differ by author (19, 40, 49, 50).
19. R. J. LaLonde, in *Means-tested transfer programs in the United States*, R. A. Moffitt, Ed. (University of Chicago Press, Chicago, IL, 2003), pp. 517-586.
20. A. L. Fernandes-Alcantara, "Vulnerable Youth: Federal Funding for Summer Job Training and Employment," (Congressional Research Service, Washington, DC, 2011).
21. A. E. Schwartz, J. Leos-Urbel, "Expanding Summer Employment Opportunities for Low-Income Youth," (Brookings Institution, The Hamilton Project, Washington, DC, 2014).
22. Youth were individually randomized within a blocked design, where blocks were defined by school and gender. All analyses control for the blocking variable. Materials and methods are available as supplementary material on Science Online.
23. E. I. Knudsen, J. J. Heckman, J. L. Cameron, J. P. Shonkoff, Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10155-10162 (2006).
24. D. M. Cutler, Are we finally winning the war on cancer? *J. Econ. Perspect.* **22**, 3-26 (2008).

25. The emphasis on those already exhibiting symptoms of the underlying problem is true of most major programs focused on youth employment. The youth elements of the National Supported Work Demonstration, JOBSTART, the National Guard ChalleNge, and the Job Training Partnership Act all focus on high school dropouts (13, 16, 17, 51). Job Corps requires applicants to need additional education, training, or vocational skills: all signs that they are already experiencing difficulties in the labor force (15). Year Up serves youth no longer in school (52). There are well-evaluated programs for high-school youth that integrate employment programming with long-term educational interventions (such as 41,42). However, because they are also lengthy and intensive – and they aim to improve educational attainment directly, which may have an independent impact on crime (29) – they do not provide a test of the hypothesis proposed here: that low-dose, low-cost employment interventions could have a larger effect than suggested by the existing employment literature when used as primary and secondary prevention.
26. R. J. Sampson, J. H. Laub, *Crime in the making: Pathways and turning points through life*. (Harvard University Press, Cambridge, MA, 1995).
27. R. Agnew, Foundation for a general strain theory of crime and delinquency. *Criminology* **30**, 47-88 (1992).
28. P. Moss, C. Tilly, "Soft" Skills and Race: An Investigation of Black Men's Employment Problems. *Work and Occupations* **23**, 252-276 (1996).
29. L. Lochner, E. Moretti, The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *Am. Econ. Rev.* **94**, 155-189 (2004).
30. J. Grogger, Market Wages and Youth Crime. *J. Lab. Econ.* **16**, 756-791 (1998).
31. B. A. Jacob, L. Lefgren, Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime. *Am. Econ. Rev.* **93**, 1560-1577 (2003).
32. See supplementary materials, section 2.7 for additional results on violent victimizations.
33. D. Deming, Better Schools, Less Crime? *Quart. J. Econ.* **126**, 2063-2115 (2011).
34. J. R. Kling, J. Ludwig, L. F. Katz, Neighborhood effects on crime for female and male Youth: Evidence from a randomized housing mobility experiment. *Quart. J. Econ.* **120**, 87-130 (2005).
35. K. A. Dodge, D. Schwartz, in *Handbook of antisocial behavior*, D. M. Stoff, J. Breiling, J. D. Maser, Eds. (John Wiley & Sons Inc, Hoboken, NJ, 1997), pp. 171-180.
36. J. Ludwig, A. Shah, "Think Before You Act: A New Approach to Preventing Youth Violence and Dropout," *Discussion Paper 2014-02* (The Hamilton Project, Brookings Institution, Washington, DC, 2014).
37. A. M. Piehl, in *The Handbook of Crime and Punishment*, M. Tonry, Ed. (Oxford University Press, Oxford, 1998), pp. 302-319.
38. S. Raphael, R. Winter-Ebmer, Identifying the Effect of Unemployment on Crime. *J. Law Econ.* **44**, 259-283 (2001).
39. The reliance on official arrest data means that violent crime is a better-measured outcome; nationally, violent crimes are 2.5 times more likely to result in an arrest than property crimes (43). Differential arrest probabilities also help explain why the control means for property and drug arrests are around half as large as for violent-crime arrests, despite the fact that non-violent crimes are more common (43). See supplementary materials, section 2.9 on how under-reporting might affect the results.
40. C. J. Heinrich, H. J. Holzer, Improving education and employment for disadvantaged young men: Proven and promising strategies. *Ann. Am. Acad. Pol. Soc. Sci.* **635**, 163 (2011).
41. J. J. Kemple, C. J. Willner, "Career academies: Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood," (MDRC, New York, 2008).
42. A. Schirm, E. Stuart, A. McKie, "The Quantum Opportunity Program Demonstration: Final Impacts," (Mathematica Policy Research, Inc., Washington, DC, 2006).
43. Federal Bureau of Investigation, "Crime in the United States 2012," *Uniform Crime Report* (U.S. Department of Justice, Washington DC, 2012).
44. L. Langton, M. Berzofsky, C. Krebs, H. Smiley-McDonald, "Victimizations Not Reported to the Police, 2006-2010," (Bureau of Justice Statistics, U.S. Department of Justice, Washington, DC, 2012).
45. S. Heller, H. A. Pollack, R. Ander, J. Ludwig, "Preventing Youth Violence and Dropout: A Randomized Field Experiment," *NBER Working Paper No. 19014* (National Bureau of Economic Research, Cambridge, MA, 2013).
46. P. Gottschalk, Can Work Alter Welfare Recipients' Beliefs? *J. Policy Anal. Manage.* **24**, 485-498 (2005)10.2307/30164060).
47. J. B. Grossman, J. P. Tierney, Does mentoring work?: An impact study of the Big Brothers Big Sisters program. *Evaluation review* **22**, 403-426 (1998).

48. R. Barnoski, S. Aos, "Outcome evaluation of Washington State's research-based programs for juvenile offenders," (Washington State Institute for Public Policy, Olympia, WA, 2004).
49. S. Raphael, in *Controlling Crime: Strategies and Tradeoffs*, P. J. Cook, J. Ludwig, J. McCrary, Eds. (University of Chicago Press, Chicago, 2011), pp. 521-565.
50. J. J. Heckman, A. B. Krueger, *Inequality in America: what role for human capital policies?*, (MIT press, Cambridge, MA, 2003).
51. Manpower Demonstration Research Corporation, "Summary and findings of the National Supported Work demonstration," (Ballinger Publishing Company, Cambridge, MA, 1980).
52. A. Roder, M. Elliott, "A Promising Start: Year Up's Initial Impacts on Low-Income Young Adults' Careers," (Economic Mobility Corporation, New York, 2011).
53. P. Z. Schochet, J. Burghardt, S. Glazerman, "National Job Corps Study: The impacts of Job Corps on participants' employment and related outcomes," (Mathematica Policy Research, Inc., Washington, DC, 2001).
54. Chicago Public Schools, in *School Data*, <http://www.cps.edu/SchoolData/Pages/SchoolData.aspx>. (2012).
55. Center for Labor Market Studies, "The Depression in the Teen Labor Market in Illinois in Recent Years," *Center for Labor Market Studies Publications, Paper 49* (2012).
56. J. D. Angrist, G. W. Imbens, D. B. Rubin, Identification of causal effects using instrumental variables. *J. Amer. Statistical Assoc.* **91**, 444-455 (1996).
57. G. W. Imbens, J. D. Angrist, Identification and Estimation of Local Average Treatment Effects. *Econometrica* **62**, 467-475 (1994).
58. H. S. Bloom, Accounting for No-shows in Experimental Evaluation Designs. *Evaluation Review* **8**, 225-246 (1984).
59. V. J. Hotz, S. G. Sanders, Bounding treatment effects in controlled and natural experiments subject to post-randomization treatment choice. *Population Research Center Discussion Paper Series No. 94-2*, (1994).
60. L. F. Katz, J. R. Kling, J. B. Liebman, Moving to opportunity in Boston: Early results of a randomized mobility experiment. *Quart. J. Econ.* **116**, 607-654 (2001).
61. A. C. Cameron, J. B. Gelbach, D. L. Miller, Bootstrap-based improvements for inference with clustered errors. *Rev. Econ. Stat.* **90**, 414-427 (2008).
62. J. M. Wooldridge, in *Handbook of applied econometrics Volume II: Microeconomics*, M. H. Pesaran, P. Schmidt, Eds. (Blackwell, Oxford, UK, 1997), pp. 352-406.
63. P. H. Westfall, S. S. Young, *Resampling-based multiple testing: Examples and methods for p-value adjustment*. (Wiley-Interscience, New York, 1993).
64. 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. Amer. Statistical Assoc.* **103**, 1481-1495 (2008).
65. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* **57**, 289-300 (1995).
66. Y. Benjamini, A. M. Krieger, D. Yekutieli, Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491-507 (2006).
67. J. J. Heckman, R. J. Lalonde, J. A. Smith, in *Handbook of Labor Economics*, C. A. Orley, C. David, Eds. (Elsevier, 1999), vol. Volume 3, Part A, pp. 1865-2097.
68. C. T. King, C. Heinrich, paper presented at the Association for Public Policy Analysis and Management's Fall Research Conference, Washington DC, 3-5 November 2011.
69. R. J. LaLonde, The promise of public sector-sponsored training programs. *J. Econ. Perspect.* **9**, 149-168 (1995).
70. P. H. Rossi, R. A. Berk, K. J. Lenihan, *Money, work, and crime: Experimental evidence*. Quantitative Studies in Social Relations (Academic Press, New York, 1980).
71. H. S. Bloom, J. A. Riccio, N. Verma, J. Walter, "Promoting Work in Public Housing. The Effectiveness of Jobs-Plus. Final Report," (Manpower Demonstration Research Corporation (MDRC), New York, 2005).
72. J. A. Riccio, "Sustained Earnings Gains for Residents in a Public Housing Jobs Program: Seven-Year Findings from the Jobs-Plus Demonstration," *MDRC Policy Brief* (MDRC, New York, 2010).
73. S. Maguire, J. Freely, C. Clymer, M. Conway, D. Schwartz, "Tuning In to Local Labor Markets: Findings from the sectoral employment impact study," (Public/Private Ventures, Philadelphia, PA, 2010).
74. C. D. Mallar, C. V. D. Thornton, Transitional aid for released prisoners: Evidence from the life experiment. *J. Hum. Resour.* **13**, 208-236 (1978).

75. C. Redcross, M. Millenky, T. Rudd, V. Levshin, "More than a job: final results from the evaluation of the Center for Employment Opportunities (CEO) transitional jobs program," *OPRE Report 2011-18* (MDRC; U.S. Department of Health and Human Services, Office of Planning, Research and Evaluation (OPRE), Washington, DC, 2012).
76. G. Farkas, D. A. Smith, E. W. Stromsdorfer, The youth entitlement demonstration: Subsidized employment with a schooling requirement. *J. Hum. Resour.* **18**, 557-573 (1983).
77. R. J. LaLonde, Evaluating the econometric evaluations of training programs with experimental data. *Am. Econ. Rev.* **76**, 604-620 (1986).
78. J. B. Grossman, C. L. Sipe, "Summer Training and Education Program (STEP): Report on long-term impacts," (Public/Private Ventures, Philadelphia PA, 1992).
79. G. G. Somers, E. W. Stromsdorfer, A cost-effectiveness analysis of in-school and summer Neighborhood Youth Corps: a nationwide evaluation. *J. Hum. Resour.* **7**, 446-459 (1972).
80. G. Walker, F. Vilella-Velez, "Anatomy of a Demonstration: The Summer Training and Education Program (STEP) from Pilot through Replication and Postprogram Impacts," (Public/Private Ventures, Philadelphia, PA, 1992).
81. J. Leos-Urbel, What Is a Summer Job Worth? The Impact of Summer Youth Employment on Academic Outcomes. *J. Policy Anal. Manage.* **33**, 891-911 (2014).
82. A. E. Schwartz, J. Leos-Urbel, M. Silander, M. Wiswall, in *IESP Working Paper Series #03-14*. (2014).
83. A. Sum, M. Trubskyy, W. McHugh, "The Summer Employment Experiences and the Personal/Social Behaviors of Youth Violence Prevention Employment Program Participants and those of a Comparison Group," (Center for Labor Market Studies, Northeastern University, Boston, MA, 2013).
84. W. S. McClanahan, C. L. Sipe, T. J. Smith, "Enriching Summer Work: An Evaluation of the Summer Career Exploration Program," (Public/Private Ventures, Philadelphia, PA, 2004).
85. R. Bender, S. Lange, Adjusting for multiple testing—when and how? *J. Clin. Epidemiol.* **54**, 343-349 (2001).
86. A. Gelman, J. Hill, M. Yajima, Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness* **5**, 189-211 (2012).
87. P. Schochet, "National Job Corps Study: Characteristics of Youths Served by Jobs Corps," (Mathematica Policy Research, Inc., Washington, DC, 1998).
88. L. L. Orr, H. S. Bloom, S. H. Bell, F. Doolittle, W. Lin, G. Cave, *Does training for the disadvantaged work?: Evidence from the National JTPA study*. (The Urban Institute, Washington, DC, 1996).
89. J. L. Lauritsen, R. J. Sampson, J. H. Laub, The Link between Offending and Victimization among Adolescents. *Criminology* **29**, 265-292 (1991).
90. M. J. Hindelang, M. R. Gottfredson, J. Garofalo, *Victims of personal crime: An empirical foundation for a theory of personal victimization*. (Ballinger Publishing, Cambridge, MA, 1978).
91. P. J. Cook, The Demand and Supply of Criminal Opportunities. *Crime and Justice* **7**, 1-27 (1986)10.2307/1147515).
92. J. Bound, C. Brown, N. Mathiowetz, in *Handbook of Econometrics*, J. J. Heckman, E. Leamer, Eds. (Elsevier, 2001), vol. Volume 5, pp. 3705-3843.
93. S. McConnell, S. Glazerman, "National Job Corps Study: The Benefits and Costs of Job Corps," (US Department of Labor, Employment and Training Administration, Washington, DC, 2001).
94. M. Cohen, R. Rust, S. Steen, S. Tidd, Willingness to pay for crime control programs. *Criminology* **42**, 86-106 (2004).
95. T. R. Miller, M. A. Cohen, B. Wiersema, "Victim costs and consequences: A new look," *National Institute of Justice Research Report* (U.S. Department of Justice, National Institute of Justice, 1996).
96. C. R. Belfield, M. Nores, S. Barnett, L. Schweinhart, The High/Scope Perry Preschool Program Cost-Benefit Analysis Using Data from the Age-40 Followup. *J. Hum. Resour.* **41**, 162-190 (2006).

Acknowledgments: This project was supported by Award B139634411 from the U.S.

Department of Labor and Grant 2012-MU-FX-0002 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. The study received approval from the University of Chicago's IRB (protocol 12-1112) with secondary approval from the University of Pennsylvania's IRB (protocol 818707). I would like to acknowledge the City of Chicago and Mayor Rahm Emanuel; the work of the Chicago Department of Family Support Services, especially Evelyn Diaz, Jennifer Axelrod, Andrew Fernandez, and Jennifer Welch; and the organizations that implemented the program: Phalanx Family Services, Sinai Community Institute, Saint Sabina Employment Resource Center, SGA Youth and Family Services, and Youth Guidance. Thanks to Chicago Public Schools partners Javin Foreman, Kylie Klein, Jennifer Loudon, and Stacy Norris; to Chief Robert Tracy, Timothy Lavery, and the Chicago Police Department; to Bob Goerge and Chapin Hall; to Roseanna Ander, Stephen Coussens, Jonathan Davis, Gretchen Cusick, Meg Egan, Harold Pollack (Chicago PI), Richard Harris, Nathan Hess, Addie Métivier, and Janey Rountree for research assistance or project support; and to Kerwin Charles, Jonathan Guryan, Charles Loeffler, Jens Ludwig, John MacDonald, Emily Owens, Stephen Raudenbush, Sebastian Sotelo, and Daniel Tannenbaum for helpful comments. Replication data is posted at the University of Michigan's ICPSR data depository (<http://doi.org/10.3886/E18627V1>); see supplementary materials section 1.5 for details. All content is the responsibility of the author and does not represent the official position or policies of the Chicago Police Department, Chicago Public Schools, Chicago Department of Family and Support Services, Department of Justice, or Department of Labor.

Supplementary Materials:

www.sciencemag.org
Materials and Methods
Supplementary Text
Tables S1 to S8
References (53-96)

Fig. 1 Intent-to-treat program effect on crime by arrest type. Figure shows control group mean, regression-adjusted treatment mean, difference between the two (ITT), and 95 percent confidence interval for the difference by arrest type. Data from Chicago Police Department administrative arrest records through 16 months post-randomization. ITT calculated using ordinary least squares regression with heteroskedasticity-robust standard errors, controlling for baseline covariates and blocking variables. Percent change relative to the control mean. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

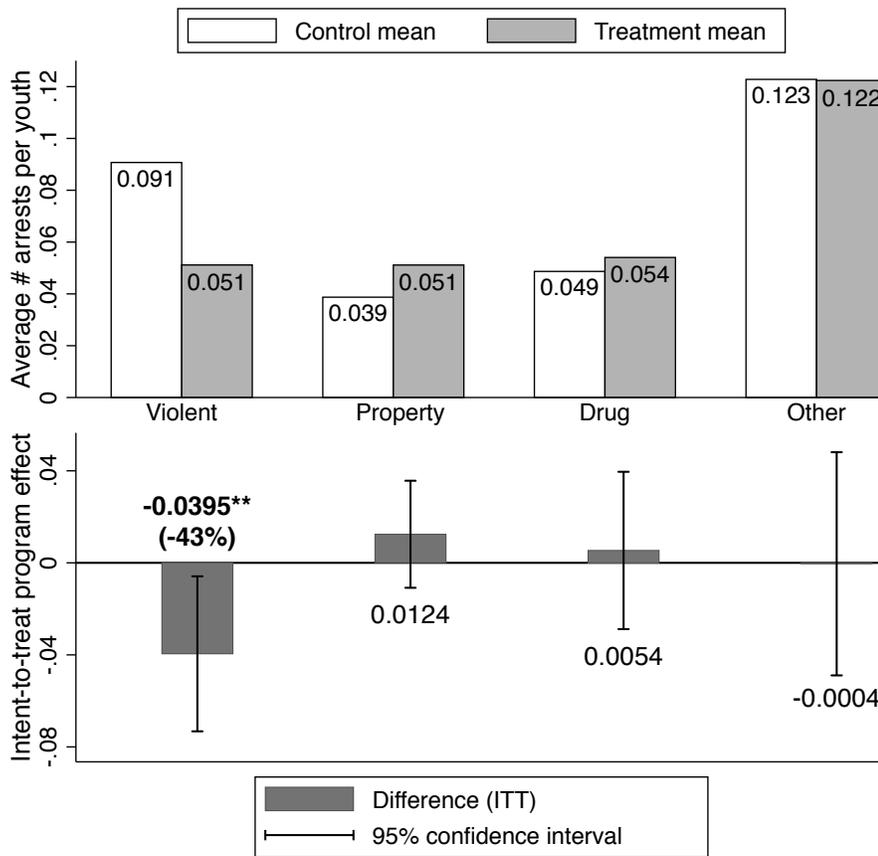


Fig. 2 Time path of violence decrease. Figure shows regression-adjusted cumulative difference in the number of violent-crime arrests between treatment and control youth by month. Random assignment lottery occurred at month 0; program ended at month 3. Regressions include baseline covariates and blocking variables. Confidence intervals calculated using heteroskedasticity-robust standard errors.

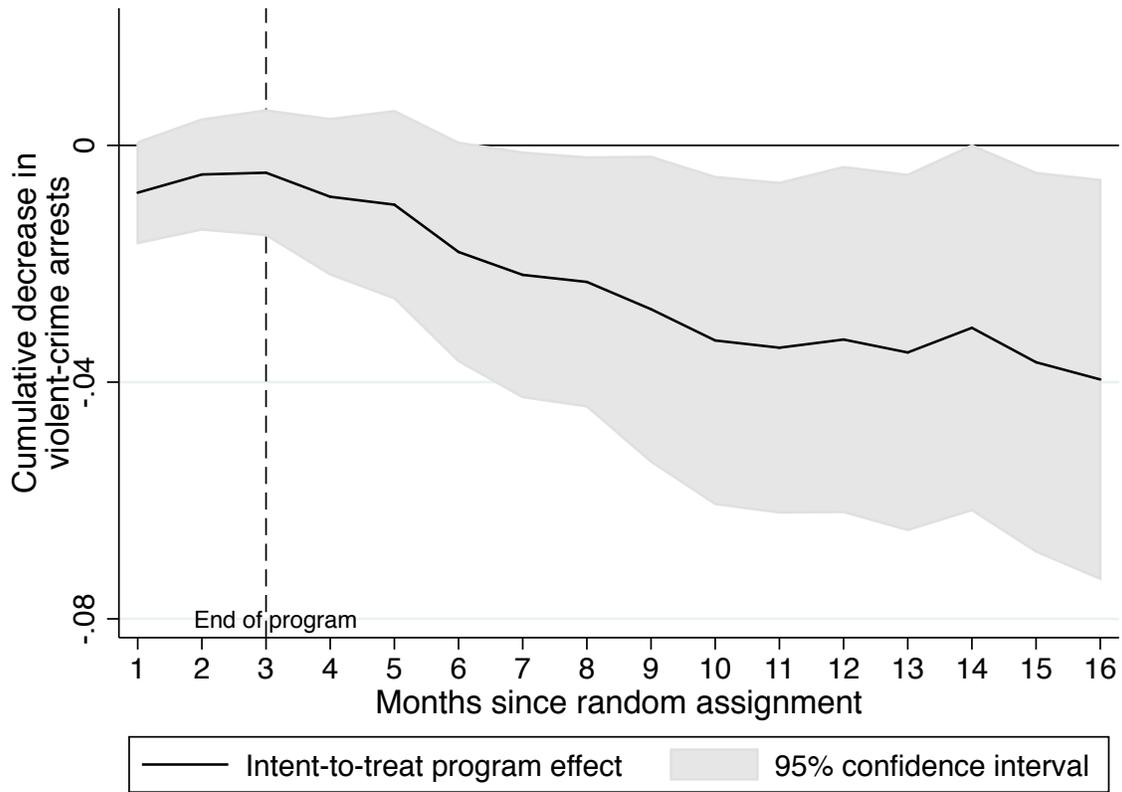


Table 1. Pre-program characteristics for treatment and control groups. Each characteristic regressed on treatment indicator and blocking variables using heteroskedasticity-robust standard errors; p-value column reports statistical significance of treatment indicator's coefficient. Demographic and schooling data from Chicago Public Schools' administrative records on pre-program year (2011-12 school year). Free/reduced price lunch is proxy for family poverty. Days absent are reported as a percentage of days enrolled (average missed = 29 days). Arrest and victimization data from Chicago Police Department administrative records. Neighborhood characteristics from 2010 American Community Survey and Chicago Police Department's community-area crime rates. "% Unemployed" is percent of civilian labor force over age 16 looking for work but without a job. Table uses non-missing data only (5 youth missing school attendance data, 60 missing fall GPA, 1 missing neighborhood violent-crime rate). Gender is not included in the table since it is captured by the blocking variables; 38 percent of the sample is male.

	Control mean (N = 904)	Treatment mean (N = 730)	P-value
Demographics			
Age	16.79	16.78	0.46
Grade	10.15	10.12	0.55
% Black	96%	94%	0.84
% Hispanic	2.9%	3.8%	0.74
% Free/reduced price lunch	92%	92%	0.98
Schooling			
% Days absent	18%	18%	0.99
GPA in fall 2011 (4 point scale)	2.37	2.32	0.93
% Enrolled in summer school 2011	8.8%	9.0%	1.00
Crime & Victimization			
% Ever arrested	19%	22%	0.37
# Violent-crime arrests	0.13	0.18	0.20
# Property-crime arrests	0.09	0.09	0.71
# Drug arrests	0.05	0.08	0.55
# Other arrests	0.15	0.19	0.55
% Ever victimized	21%	24%	0.36
# Victimitations	0.29	0.33	0.30
Neighborhood Characteristics			
% Unemployed	19%	19%	0.94
% Below poverty line	33%	35%	0.71
Median household income (US\$)	35,665	34,321	0.58
Violent crime rate (per 100,000)	2,128	2,136	0.95

Table 2. Intent-to-treat effect on violent-crime arrests by treatment group. Results from regression of number of violent-crime arrests on separate treatment group indicators, baseline covariates, and block fixed effects. Bottom row shows the magnitude and standard error of the difference in effect across treatment arms. Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Jobs Only	-0.0391* (0.0205)
Jobs + SEL	-0.0399** (0.0203)
Difference between groups	0.0008 (0.0220)

Supplementary Materials for

Summer Jobs Reduce Violence among Disadvantaged Youth

Sara Heller

correspondence to: hellersa@sas.upenn.edu

This PDF file includes:

Materials and Methods
Supplementary Text
Tables S1 to S8
References (53-96)

1. Materials and Methods

Materials

1.1 Crime Data

The main outcome data consist of arrest records from the Chicago Police Department (CPD). Data were matched to study youth using name, date of birth, and gender with a probabilistic matching software called Merge Tool Box. When recording arrests, CPD uses fingerprint technology to identify individuals. The arrest data therefore include basically every CPD arrest of an individual, even if he or she submits an alias at the time of arrest.

To separate crimes by type, I identify the most severe charge associated with each arrest. 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, and motor vehicle theft. Drug crimes include both possession and dealing, and other crimes include other kinds of offenses (e.g., trespassing, vandalism, cruelty to animals, weapons violations, parole violations, outstanding warrants, etc.). I exclude motor vehicle-related arrests because they are so infrequent in the data (only 0.6 percent of the sample had such an arrest at baseline). Any serious vehicle-related offenses like auto theft, DUI, or forged registrations are included in the main crime categories. I then count the number of pre- and post-program incidents of each type, defining “post” as after the date of the lottery, since program providers began informing youth about program slot offers immediately.

Note that the data are limited to arrests conducted by the Chicago Police Department, which only cover the city limits. Logically, it could be that treatment increases time spent outside the city and so reduces CPD arrests without actually reducing violent crime. Other research on Chicago juveniles suggests the magnitude of this kind of differential censoring is likely to be quite small. Statewide arrest data for a similar population suggests that about 5 percent of arrests for Chicago male youth living in high-crime neighborhoods occur outside CPD’s jurisdiction (45). With a control mean of 0.091 violent-crime arrests per youth, that would imply this study’s control group is arrested for about 0.005 violent crimes per youth outside of CPD’s jurisdiction. In order to explain the entire treatment effect (0.0395 fewer violent-crime arrests per youth), treatment youth would have to increase their travel and offending rate outside of the city by a factor of eight relative to the control group – but only for violent crimes, since there is no change in property, drug, or other arrest rates. All summer jobs were inside city limits, so treatment did not directly encourage out-of-city travel. Additionally, qualitative evidence collected from some of these same neighborhoods suggests that regular travel outside of Chicago is quite difficult and rare, even for adult residents (6). As a result, it seems implausible that differential censoring can explain the entire observed decrease in violence.

Administrative arrest data avoid limitations of self-reported crime like social desirability bias, which might be particularly problematic given that the treatment group received a fair amount of money from the program and so may be less willing to admit wrongdoing than the control group. Nonetheless, official arrest records are not without limitations as measures of crime and violence. They tend to understate the overall amount of crime, since many crimes do not result in an arrest, and they capture both criminal and police behavior. 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 (53).

I also present results on violent victimizations below. In order to focus on violence, I only count offenses against a person, excluding non-violent victimizations such as vandalism, forgery, etc., which are very rare in the data. The victimization data are considerably more noisy than the arrest records. Victimization are recorded when a police officer is called to an incident. CPD officers record the name and date of birth of each victim in the data, but unlike in the arrest data, do not use a single identifying number for an individual across incidents. I therefore use probabilistic matching to group incidents by individual, so that spelling does not have to match exactly across incidents to be linked to the same person. I then use the same method to match the victim data to the study sample. If an individual offered a drastically different name or date of birth across separate incidents, or if police recorded the wrong name, the matching process would not link both incidents to the study youth. The victimization data are therefore likely to understate the extent of victimization in the sample (which tends to be true of victimization data in general, since only about half of violent victimization are reported to the police nationwide (44)).¹

The victim data may also reflect changes in the willingness to call the police when victimized. If a summer job changes the propensity to report crime, the estimated treatment effect could either over- or under-state the true change in victimization (depending on if a summer job discourages or encourages reporting). Nonetheless, since each violent incident has the same social cost regardless of whether a study youth is the victim or offender, it is useful to have some measure, even an imperfect one, of victimization as well.

It is also theoretically possible that different study youth were both a victim and an offender in a single incident, which would mean adding the victim data to the arrest data could double-count that incident. Although it is difficult to decisively link an incident in the arrest data with the same incident in the victim data, both data sets have date/time fields and incident descriptions. Comparing the two, only 6 post-lottery violent-crime arrests share a date and hour with observations in the victimization data, 2 of which have descriptions that do not appear to match. The remaining 4 may not have been the same incident, but they are arrest/victimization pairs that occurred at about the same time (and are evenly split between treatment and control offenders). As such, double counting seems unlikely to be a major issue.

1.2 Schooling Data

Youth included their Chicago Public Schools student ID numbers on their applications, which were cross-checked with CPS records and matched directly to schooling records using these unique identifiers. Baseline characteristics come from the 2011-12 academic year. Because the lottery occurred 2.5 weeks before the end of the school year, data for the entire year may include a very short period of time when outcomes could have been affected by randomization status: Employment providers started informing youth they had been selected just before the end of the school year. Wherever possible, I exclude this data from the baseline characteristics (i.e., I use

¹ Importantly, 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. If name-based misreporting were correlated with a determinant of treatment heterogeneity, however (e.g., more pro-social youth ensure their names are spelled correctly when victimized and benefit more from the program), the victim data would only represent the treatment effect for those well represented in the victim data.

only fall grades, and I only count disciplinary incidents through the date of the lottery). The only exception is in the baseline attendance data, which is only available for the entire year. Although it is possible that annual 2011-12 attendance captures an outcome, not just a baseline characteristic, this change is likely to be quite small. Indeed, absences are balanced across treatment and control groups (p-value of difference = 0.99).

Schooling outcome data are from the 2012-13 school year. The majority of students in 12th grade prior to the program will not have data for this school year (unless they fail to graduate, which is relatively rare conditional on staying in school through 12th grade). So schooling outcome analyses use youth in 8th – 11th grade during the pre-program year only.

There are 40 youth who do not appear in the CPS attendance file for the 2012-13 school year and 154 with no information on grades earned. (Youth might have positive attendance but no grades because they only attended a few days at the beginning of the school year, or because of data entry errors in either file.) Given that only about half of students who enter the study high schools graduate within 5 years during this period (54), it would not be uncommon for students to have missing data because they stopped attending school. However, it is also possible that some of these students transferred to other school districts and so have schooling outcomes that are not captured in the CPS data. Although the district records whether students transfer (28 are marked as transfers), school records on dropouts and transfers are not always accurate: schools face no penalty for transfers but do for dropouts and so have an incentive to misreport.

As a result, I handle the missing schooling data in a variety of ways to ensure that results are not sensitive to the decision. I start by reporting results for only those who appear in the data, which implicitly assumes that missing data are missing completely at random (i.e., the missingness mechanism is uncorrelated with students' observable and unobservable characteristics). Since this is a strong assumption, I then use logical imputation by assuming that anyone not in school has dropped out and assigning zeros for days present and GPA. Lastly, I make use of CPS's records on transfers and impute group means (treatment or control) for the students who CPS reports have transferred (which assumes their data is missing completely at random) and assign 0s otherwise. Since only 2.9 percent of the 8th – 11th grade sample has missing attendance data, and since the same proportion of data is missing for treatment and controls (treatment-control difference in missingness adjusting for blocking variables = -0.006, $p = 0.52$), the different ways of handling the missing data do not make much substantive difference. (The same is true for academic grades, where treatment-control difference in missingness adjusting for blocking variables = -0.003, $p = 0.87$.)

Methods

1.3 Experimental Design

The program was administered by Chicago's Department of Family and Support Services (DFSS) and implemented by five non-profit community organizations. Because DFSS designed OSP primarily as a violence-reduction intervention, program operators focused on recruiting a population of youth at high risk of violence involvement in the city. To identify this population, Chapin Hall – a partner research organization and data warehouse – used 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. With models designed to assess the risk of violence, Chapin Hall identified the 5,000 youth at highest risk in the city. The City selected the 13 high schools with the highest numbers of these students for OSP. All

students currently enrolled in the 13 program schools, or planning to enroll as rising 9th graders in the fall, were eligible for the program as long as they would be between ages 14 and 21 at the program's start.²

The schools, community organizations, and DFSS partnered to recruit 1,634 eligible youth for 700 available slots using both online and paper applications. In May 2012, I randomly assigned each applicant to one of three groups: 350 to the jobs-only treatment group, 350 to the jobs + SEL group, and the remaining applicants to the control group.

Randomization occurred within school-gender blocks for three reasons. First, each employment provider was assigned to work with a certain number of schools. School blocking ensured that each provider served roughly the same number of youth, despite different numbers of applications across schools. Second, application rates were lower among males, but DFSS wanted to include more males in the program given their disproportionate involvement in violence. Blocking on gender allowed the over-selection of males. Third, since the outcomes of interest vary by school and gender, the blocks contribute to the power of the study by explaining residual variation in the outcomes.

The control group was randomly ordered to form a waiting list (within school-gender blocks by treatment arm). Before the program began, 30 treatment youth (4.3%) declined to participate. Thirty youth who were originally in the control group were therefore taken off the randomized wait list (within the same school-gender-treatment groups as the decliners) and shifted to the treatment group. Because these youth were randomly selected, the analysis includes them as part of the treatment group. Those who declined to participate are also included to preserve the integrity of random assignment. The rest of the control group was strictly prohibited from participating in OSP, although they were free to pursue other employment opportunities or summer programming. Prior research suggests that the control group's employment rate is likely to be quite low; low-income black teenagers in Illinois had an employment rate of 9 percent in 2010 (55).

As discussed in the main text, randomization successfully balanced all observable characteristics across treatment and control groups.³ Since treatment youth were randomly assigned to two groups, I can also test the baseline equivalence of each treatment arm with the control group separately. I find no more significant differences between each group and the control group than would be expected by chance.⁴

² The racial and ethnic composition of the sample – almost entirely African-American – was somewhat surprising given that several schools on the largely Hispanic West Side were included. One possibility is that legal work was less attractive to students who have a higher likelihood of being (or having family) ineligible to work legally in the U.S.

³ The test for joint significance of all baseline variables uses additional variables not shown in Table 1, including: free/reduced price lunch, age, black and Hispanic dummies, grade dummies, percent of enrolled days absent, the number of each grade category received in fall 2011 (A through F), a dummy for attending summer school in 2011, the total number of minor and severe disciplinary incidents, the number of days of in- and out- of school suspensions, neighborhood characteristics (unemployment, median income, fraction below the poverty line, fraction of adults with at least a high school degree, and violent crime rates), and the number of each type of arrest (violent, property, drug, motor vehicle, and other) and victimization (violent and non-violent).

⁴ Over 42 baseline variables, there are 86 different hypothesis tests (one for each treatment group for each variable). With this many tests, one would expect between four and five treatment group-control differences to be significant at the five percent level and about 9 at the 10 percent level purely due to chance. The actual imbalance is less than expected: the jobs-only group has more baseline C's during the fall semester before the program ($\beta = 0.19$, $p =$

1.4 Analysis Methods

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

$$(1) \quad Y_{ibt} = T_{ib}\pi_1 + X_{ib(t-1)}\beta_1 + \gamma_b + \varepsilon_{ibt1}$$

where T_{ib} is an indicator for being randomly assigned to the treatment condition, $X_{ib(t-1)}$ are observed variables from administrative records measured at or before baseline, γ_b are block fixed effects, and ε_{ibt1} 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.⁵

The parameter π_1 captures the intent-to-treat (ITT) effect of being assigned to the program group. As the ITT measures the impact of offering services on the outcome Y , it may be a policy-relevant parameter in this case. For example, a school principal deciding whether to market the program in her school may be interested in how much it will reduce violence among all the applicants, not just among participants.

Nonetheless, because not all youth end up participating, the ITT will understate the effects of actually participating in the program for those youth who choose to participate. Under the typical relevance and exogeneity assumptions for instrumental variables (56)⁶, this latter set of effects can be recovered from the experimental data. I perform this estimation through a two-stage least squares strategy, in which random assignment (T_{ib}) is an instrument for actual participation (P_{ibt}), and P'_{ibt} is the predicted probability of participation from equation (2) (57, 58):

$$(2) \quad P_{ibt} = T_{ib}\pi_2 + X_{ib(t-1)}\beta_2 + \gamma_{b2} + \varepsilon_{ibt2}$$

$$(3) \quad Y_{ibt} = P'_{ibt}\pi_3 + X_{ib(t-1)}\beta_3 + \gamma_{b3} + \varepsilon_{ibt3}$$

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. If, as is perhaps more likely, treatment effects are heterogeneous across youth, then the coefficient π_3 estimates a local average treatment effect – the effect of participation on those who comply with random assignment.⁷ Because there is no control crossover (no always-takers) in this setting, π_3 provides an estimate of the treatment-on-

0.054, CM = 1.95) and more pre-program violent victimizations ($\beta = 0.07$, $p = 0.079$, CM = 0.25). No other differences are significant below the 10 percent level. While the differences are entirely consistent with chance variation rather than randomization failure, none of the outcome results are sensitive to controlling for the baseline differences.

⁵ Baseline covariates included in the main regressions are: indicators for free or reduced price lunch, being Hispanic, grade of enrollment, summer school 2011 attendance, and missing baseline data (0s imputed for missing variables), as well as age at program start, the number of each grade type (A through F) in fall 2011, number of in- and out-of-school suspensions, number of minor and major disciplinary incidents, percent of days enrolled absent, number of each type of arrest (violent, property, drug, motor vehicle, and other) and victimization (violent and non-violent), and neighborhood characteristics (percent unemployed, below the poverty line, and with a high school degree, median income, and violent crime rate). None of the substantive conclusions are different if these variables are excluded from the outcome regressions, but the covariates do improve precision.

⁶ In order for the random assignment variable, T_{ib} , to be a valid instrument, it must be correlated with program participation, P_{ibt} , and uncorrelated with ε_{ibt3} .

⁷ When treatment effects are heterogeneous, T_{ib} , 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.

the-treated (TOT): the effect of the program on those who choose to participate.⁸ To help judge the magnitude of the TOT estimates, I also estimate the average outcomes of those youth in the control group who would have participated had they been assigned to treatment, or the “control complier mean” (CCM) (see 60).

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. One assumption that might fail in this case is the independence of observations, especially within schools (since youth are more likely to share social networks and act in groups with other students in their schools). With only 13 schools, the asymptotics on which a standard clustering adjustment relies are not valid. I instead check the robustness of the results to non-independence within schools by using the wild-t bootstrap, which Cameron, Gelbach & Miller (61) show performs well with a small number of clusters.

I also 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 (62). I use Huber-White robust standard errors to allow for over-dispersion, relaxing the Poisson distributional constraint that the mean equals the variance. To ensure convergence, I use a limited set of baseline covariates.⁹

Lastly, as in any experiment testing program effects on multiple outcomes, one might worry that the probability of Type I error increases with the number of tests conducted. In addition to pre-specifying a small number of focal hypotheses, I also account for multiple inference concerns in two different ways. First, I adjust p-values using a free-step down resampling method,¹⁰ which controls the family-wise error rate (FWER), or the probability that at least one of the true null hypotheses in a family of hypothesis tests is rejected (63, 64). The FWER approach is useful for controlling the probability of making any Type I error, but it trades off power for this control. An alternative is to control the probability that a null rejection is a Type I error (the false discovery rate, or FDR), increasing the power of individual hypothesis tests in exchange for allowing some specified proportion of rejections to be false (65, 66). I use both methods to assess the results’ robustness to adjustments for multiple hypothesis testing.

⁸ I also estimate the effect of the two treatment arms separately by including two separate random-assignment indicators. As Hotz and Sanders (59) note, with more than one treatment arm, the TOT is not necessarily point identified without additional assumptions. The basic intuition is that the composition of compliers could vary across the two treatment arms if the type of treatment offered affects participation decisions. One could bound the TOT as Hotz and Sanders suggest, or interpret the separate IV estimates as a local average treatment effect rather than a TOT. For simplicity, I focus on the ITT when comparing the two treatment arms. The IV results by treatment arm do not change any substantive conclusions.

⁹ I also use Stata’s “difficult” option to change the stepping algorithm, as the drug-crime likelihood function exhibits too much non-concavity to converge otherwise.

¹⁰ 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 (63, 64). See (45) for a more detailed description.

1.5 Replication Data Set

Data and Stata code for replication are posted at the University of Michigan's ICPSR data depository (<http://doi.org/10.3886/E18627V1>). It is important to note that the public data set is not the exact data used in the analyses reported here. I have transformed some of the variables in order to comply with restrictions set by the agencies that own the data and to better de-identify the data. The transformed variables include all information from the Chicago Public Schools (demographic and academic variables), as well as all neighborhood characteristics and cost of crime variables.

The new variables are equal to the original variable minus the block mean (as described above, youth were blocked by school and gender, then individually randomly assigned within blocks). This transformation obscures individuals' personal information but preserves the ability to replicate the results in the paper. The reason is that all regressions reported here include block dummy variables (or block "fixed effects"). As the theory of partial regression explains, including fixed effects generates the same point estimates as running a regression with all de-meaned variables. Because the replication regressions also include covariates (and in some cases, dependent variables) that are not de-meaned, they still include block fixed effects. The degrees of freedom and standard errors therefore remain correct when using the de-identified data.

The main implication of the transformation is that none of the group means can be recovered from the replication data for the de-identified variables. Specifically, the group means in Table 1 for demographics, schooling, and neighborhood characteristics will not match the means from the replication data. The same is true for the control means and control complier means in Table S7. The point estimates and standard errors for all regression results, however, are replicable, as are all means for the main arrest variables. See the "ReadMe.txt" file that accompanies the replication data for further details.

2. Supplementary Text

2.1 Prior Literature on Employment Programs

The extensive evidence on employment programs has been reviewed elsewhere (19, 67-69). This section does not attempt to provide a comprehensive summary of this literature, but rather to explicate several claims about particular aspects of the evidence made in the main text.

A. The evidence on public employment programs among adults is mixed.

Although the main text focuses on how crime responds to employment interventions, employment and earnings outcomes also show mixed results among adults, stemming partly from heterogeneous effects by subgroups. Adult women – the demographic group least involved in violence – show the most consistent employment improvements, while the evidence is more mixed for adult men and tends to vary by population, intervention strategy, and duration (17, 51, 70-73).

Though crime impacts tend to receive less attention in the literature, they also vary among adults. One small-scale intervention with intensive case management reduced crime overall (74), but a scaled-up version did not (70). Transitional jobs programs for ex-offenders, though they have not shown crime effects overall, do reduce recidivism among certain subgroups (14, 75). Raphael (49) reviews the evidence on adult employment programs and crime, concluding that while some employment-based

intervention strategies may reduce crime among adults, results vary enough across studies to be hesitant about positive conclusions.

B. The only successful employment programs for youth have been intensive and expensive.

Here again, the main text focuses on crime outcomes for youth, though this is also true for employment outcomes. The large-scale experimental studies of programs for youth that are typically considered employment-based, such as the National Supported Work Demonstration, JOBSTART, and the Job Training Partnership Act (JTPA), show no improvement in youth employment (*16, 17, 51*). The most commonly cited success for youth employment is Job Corps, a mostly residential program that averages 8 months of participation and costs over \$25,000 per youth in 2014 dollars (*15, 53*).

As discussed in the notes to the main text, other interventions that combine a focus on education with various types of career services and job training are also sometimes considered employment programs. Programs like the National Guard ChalleNGe, Year Up, and Career Academies have been shown to improve later employment outcomes (*13, 41, 52*), although a variation of the latter was less successful (*42*). These programs are also very intensive and expensive: the ChalleNGe involves 2 weeks of physically-demanding activity, a 20-week residential stay, and a one-year non-residential period (over \$17,000 per youth in 2014 dollars); Career Academies span a 3 or 4-year high school period (cost not reported); and Year Up involves 6 months of paid training and coursework followed by a 6-month paid internship (cost not reported).

In terms of crime outcomes, only Job Corps and JOBSTART, which offers education and skills training for an average of 6.8 months and costs \$9,800 per youth in 2014 dollars (*16*), appear to reduce arrests among youth. (Although in the JOBSTART study, all crime outcomes are self-reported. So one might worry that treatment youth were just less willing to admit wrong-doing while in the program). When JTPA attempted to deliver similar services in a less intensive (and less expensive) way, arrests actually increased (*17*).

One possible reason that crime declines occur only in the most intensive programs is that lengthy periods of supervision during the program create an “incapacitation” effect rather than internal behavioral change. This possibility is consistent with the fact that in both studies, crime decreases fade out quickly after the time when most youth leave the programs. In Job Corps, non-residential males, who had far less supervision than their residential counterparts, showed no decrease in crime – also consistent with the idea that the content of the programming itself may have been less important than the incapacitation effect from the residential stay. If part of the reason crime among out-of-school youth only falls in intensive programs is because of increased supervision, the same results might be achieved at a significantly lower cost than a full-scale employment program requires.

C. Summer jobs programs are widespread but poorly researched.

The federal government began funding youth summer jobs programs in 1964. In 1998, the Workforce Investment Act moved specific funding for summer jobs into the more general Youth Activities budget but still required that they be part of funded agencies’ offering to low-income youth (*20*). As a result, many cities have long-standing summer

jobs programs, including Washington D.C., Boston, New York City, Minneapolis, and St. Louis.

In his review of job training and employment programs, Lalonde (19) writes that despite serving between 489,000 to 790,000 participants annually in the 1980s and '90s, the summer youth program funded by the Job Training and Partnership Act “received relatively little attention from program evaluators,” noting only one non-experimental evaluation, the Youth Incentive Entitlement Pilot Project (76), which matched 4 treatment communities to 4 comparison communities to estimate program effects (p. 532). In addition to power challenges, it is not clear that this kind of matched comparison provides a convincing estimate of the program’s causal effects (67, 77). Another review also determines that the evidence is “not sufficient for drawing conclusions about the effectiveness of summer jobs” (20, p. 26) .

In particular, a few studies on programs carried out from the 1960s to 1980s, which included summer jobs along with a range of non-summer services, find some positive effects on schooling or earnings, especially for black males (76, 78, 79). Yet all but one use non-experimental designs susceptible to selection bias, and the single experiment (78, 80) provides both treatment and control groups with training and employment, identifying only the effect of an additional life skills and sex education curriculum which was not offered to the control group.

A single recent study provides convincing evidence on summer jobs and schooling from a natural experiment in New York City, finding small increases in attendance and test-taking for those attending school, especially among those with low pre-program attendance (81). A follow-up study on treatment heterogeneity is in progress (82). In terms of crime, there is a promising suggestion that summer jobs decrease delinquency (83). But a reliance on non-experimental comparisons as well as self-reported outcome data means that any observed behavioral differences could be because participants are more hesitant than the comparison group to report wrong-doing, or from selection into the program.¹¹

One study of a Philadelphia-area program similar to One Summer Plus claims to be a randomized controlled trial that would, in theory, overcome the concern of selection bias (84). Although the authors do not show any descriptive statistics or results on crime, nor describe the crime data used, they do report that they matched study youth to criminal records and found that “very few” of the study youth had been arrested, so there was no treatment effect on crime (p. 18). While a program serving youth who are not generally criminally-involved may not be the strongest test of whether summer jobs can reduce crime, the study suffers from a larger issue. Despite the use of random assignment to allocate program slots, the analyses reported in the paper appear to be entirely non-experimental, only comparing participants to non-participants.¹² Unfortunately, this

¹¹ The treatment and comparison groups were quite different on observable characteristics in this study. For example, 48 percent of the treatment group was female versus 37 percent in the comparison group, and half of treatment youth were African-American, compared to 35 percent of the comparison group. The extent of apparent selection into the program makes it difficult to be confident that reductions in self-reported delinquent behavior are due to the program itself rather than other, unobserved differences between the groups.

¹² The methodology appendix explains that the control youth who actually received treatment were dropped from the analysis, and the program effects were estimated by including an indicator variable for participation – not random

means that the estimates in the paper are subject to the same kind of selection bias concerns as if random assignment did not take place. As such, it does not estimate the causal effects of a summer jobs program.

A recent review argues that the available evidence on summer jobs suggests they have positive effects, especially on schooling (21). But as this section explains, most of the studies (with one exception) cannot separate selection into the program from the program's effect, and none report convincing estimates of effects on crime and violence. The author could not locate any studies with stronger research designs on summer jobs and crime that isolate their causal effect.

2.2 Program Description

One Summer Plus is structured much like summer jobs programs across the country: a local government agency partners with community and non-profit groups to provide city-funded summer jobs to youth. In this case, Chicago's Department of Family and Support Services targeted disadvantaged youth at high risk of violence involvement, who were served by five community-based non-profit organizations on the south and west sides of Chicago (Sinai Community Institute, SGA Youth and Family Services, St. Sabina Employment Resource Center, Phalanx Family Services, and Youth Guidance). Youth received one of two versions of the program for eight summer weeks¹³: jobs only or jobs + social-emotional learning (SEL, described further below).

In both versions of OSP, youth are offered 5 hours per day, 5 days per week of programming (jobs-only youth work for all 5 daily hours; jobs + SEL youth work for 3 and participate in SEL for 2). They earn Illinois minimum wage (\$8.25/hour) and receive one meal per day, plus bus passes when appropriate. The service providers locate jobs with the goals of aiding youth in exploring career interests and aptitudes, building vocational knowledge, developing team and leadership skills, and practicing creative thinking and problem-solving. Because of restrictions imposed by a private funder in the study year, jobs were only in the non-profit and government sectors (although this restriction was relaxed in later years). Jobs included positions as summer camp counselors, workers in a community garden, YMCA office and activity staff, office assistants for an alderman, etc. Youth were also assigned job mentors – adults who helped youth learn to be successful employees and to navigate barriers to employment (transportation, family responsibilities, conflicts with supervisors, etc.) – at a ratio of about 10:1.

One hypothesis for why prior youth employment programs require high intensity to succeed is that disadvantaged adolescents may lack the “soft skills” to engage with less intensive pro-social programming (40). To test one strategy for addressing this issue, half of the treatment group was randomly selected to attend SEL sessions. The motivating idea for the SEL programming is to

assignment – in a regression analysis. This effectively compares actual participants to those who did not participate (both controls and treatment non-participants), thus re-introducing selection bias concerns.

¹³ The program was originally designed to run over 7 summer weeks, but additional funding allowed for an optional week-long extension of the jobs component. Eight weeks of programming were offered but not required, and in the 8th week there was no SEL programming. One service provider also offered access to additional, optional programming outside of OSP (like drama, graphic design, and fitness activities), but these activities were not funded by the program. Although the study cannot separate out the effect of these additional activities, there is very little empirical evidence that such extra-curricular programming could be the sole mechanism driving the observed violence reduction.

help youth learn to understand and manage the aspects of their emotions and behavior that might interfere with successful participation and employment (e.g., the inclination, not uncommon among adolescents, to snap defensively at a someone offering constructive criticism).

SEL sessions focus on emotional and conflict management, social information processing, and goal setting. The curriculum differed somewhat across the two providers, but both are based on a manualized curriculum guided by cognitive behavioral therapy principles, which focus on helping youth to track how their thoughts and beliefs lead to actions, and how to better control that process. Prior research has shown that similar programming can reduce violent crime and create lasting improvements in school engagement on its own (45); its inclusion in OSP was to test whether, in combination with employment, it could increase program participation and improve outcomes more than the jobs alone.

2.3 Program Participation

Daily program attendance data was tracked by the non-profit organizations implementing the program. Table S1 shows detailed participation information. Overall, three-quarters of the youth who were offered the program actually participated. Almost every participant was placed in a job, with the exception of a few youth recorded as having SEL hours but whose job records were apparently lost: Program providers reported that an additional 28 participants were placed in jobs, but no employment records are available for these youth.

The lack of official employment records on these 28 youth explains why jobs placement is not 100 percent among participants, which seems likely to be due to recording errors rather than participation in SEL without a job. This may not be the only source of measurement error; program providers also did not consistently track hours spent in SEL versus jobs, and occasionally included the extra hours that participants chose to stay at their worksites that were not part of their paid work. Providers did have a strong incentive to record who participated at all, however, since the City held them accountable for filling a particular number of slots. As such, I focus on measuring participation on the extensive margin rather than using the type of activity or number of hours worked.

Table S1 shows that of those who started the program, about 90 percent finished at least the original 7 weeks of the program. After the program began, administrators added an optional extension week, so the final program spanned 8 weeks. The table also shows participation rates by treatment arm. None of the differences in participation across the jobs-only and jobs + SEL youth is statistically significant (with the exception of SEL participation, by design).¹⁴ There was a small amount of treatment-arm crossover: 6 percent of the jobs-only group also participated in SEL. Because program providers did not have contact information for the control group, there is no control crossover (although the control group was free to pursue other services not related to this particular program).

2.4 Treatment on the Treated

Table S2 shows the effect of participation for those who choose to participate (TOT). For that group, violence decreases by 0.053 arrests per youth ($p = 0.019$, control complier mean = 0.121). The fact that the control complier mean is higher than the overall control mean suggests that

¹⁴ Youth in the jobs + SEL treatment arms participated in slightly more days than the jobs-only youth (25 versus 27). This may be some indication that SEL did increase program participation, although the difference is just above traditional levels of statistical significance ($p = 0.12$).

there was some negative selection into who participated in the program. This is consistent with information from providers, who reported that the youth who declined to participate often said they had already committed to other jobs or summer activities.

2.5 Robustness Checks

All violence results are robust to allowing for non-independence in outcomes within schools and using non-linear specifications. As discussed in the methods section, to allow for correlated outcomes among youth who attend the same schools, I use the wild-t bootstrap rather than clustered standard errors due to the small number of clusters. Allowing for correlation in outcomes within schools (measured at the time of application), the wild-t p-value for violent-crime arrests is 0.042. In other words, the substantive conclusion of a significant decrease in violence does not change.

Since all the crime outcomes are counts of post-program arrests, I also use Poisson regression with robust standard errors to allow for over-dispersion. As shown in Table S3, the average marginal effects are extremely similar to the ITT estimates, with the implied percent change in violent-crime arrests (percent change = $[e^\beta - 1] * 100$) just slightly smaller than for the OLS estimate (37 percent). The conclusions are not sensitive to functional form.

Given that only violence shows a statistically significant change across all the criminal behaviors measured, one might be concerned that the violence results are due to chance (i.e., Type I error). The extent to which the p-value on violent-crime arrests should be adjusted in this case, given that violence was the single outcome of most interest prior to the analysis, and the best way to make the adjustment, is a matter of debate in the literature (85, 86).

Prior to performing the analysis, I pre-specified the primary outcome measures of interest as arrests for (in order) violent, property, drug, and other crime. So one might argue for no p-value adjustment, since violence was always the first and most important outcome measure of interest. Nonetheless, the decrease in violent-crime arrests is robust to adjustments for the number of hypothesis tests involved in the analysis of this set of primary outcomes.

For the four primary outcomes, I calculate adjusted p-values controlling for the family-wise error rate (63, 64). The adjusted p-value for the treatment impact on violent-crime arrests is $p = 0.08$. 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. An alternative adjustment, which allows a specified probability of Type I error, controls the false discovery rate or FDR (65, 66). Allowing for 10 percent of null hypothesis rejections to be false across the 4 crime outcomes (i.e., setting the FDR's p-value counterpart, q , to 0.1), I also reject the null hypothesis of no program effects for violent-crime arrests. Overall, the decrease in violence is robust to adjustments for multiple hypothesis testing.

It is also possible to adjust the p-values for the sets of hypothesis tests that were pre-specified as of secondary interest (e.g., the difference between treatment arms and the schooling outcomes presented below). Given that none of those results is statistically significant, however, any adjustment will not change the substantive conclusion of the analysis.

It is also worth noting that the tests of heterogeneity discussed below were not included in the pre-specified sets of outcomes and are treated as exploratory. This is because it was clear prior to the analysis that, given the size of the subgroups, the study was not powered to detect even fairly large group differences.

2.6 Other Crime Impacts by Treatment Arm

As with violent crime (Table 2), none of the other crime differences across treatment arms is statistically significant, and most point estimates are quite small (Table S4). However, power is a bit more of a concern for other crimes (e.g., trespassing, vandalism, etc.), where the point estimates are in opposite directions with non-trivial magnitudes of about 20 percent of the control mean (p-value of difference = 0.281).

2.7 Heterogeneous Treatment Effects

Heterogeneous treatment effects are also of interest. Table S5 separates program effect estimates for three different subgroups of youth defined during the pre-program year: males versus females, 12th graders versus those still likely to be in school during the follow-up year, and those with versus without a prior arrest for a violent crime. These exploratory results should be interpreted with caution, both because of the number of hypothesis tests involved and because the study was not powered to detect these sorts of subgroup differences.

Indeed, the interactions are too imprecisely estimated to tell a clear story. Overall, the point estimates suggest that the program may prevent more violent-crime arrests for males, but the large standard errors cannot rule out the absence of gender differences (or large differences).¹⁵ Similarly, students in 12th grade the year before the program, who are more likely to be entering the labor force than returning to school, show a bigger decrease in violence relative to those who remain school-age (who show significant decreases on their own). But the interaction is too imprecisely estimated for any strong conclusion.

Youth with and without a prior violent-crime arrest both show a similar proportional decrease in violent-crime arrests (about 40 percent). But the point estimate on the interaction suggests that the number of crimes prevented is more than four times as large for those who already have a history of violent offending (a decrease of 0.127 versus 0.029 arrests, although the difference is imprecisely estimated). This is because those with a violent history have a much higher risk of future violence involvement (control group members with prior violent-crime arrests have an average of 0.321 violent-crime arrests after randomization, compared to only 0.068 for those without a baseline violent-crime arrest). In other words, there is more room for prevention among those with the highest rates of violence. Using prior violent-crime arrests as a way to identify the youth at the highest risk of violence may be one way to maximize the program's impact on the overall level of societal violence.¹⁶

At first, the fact that those already involved in violence may respond more to treatment could seem to contradict the study's motivating idea that prevention is easier and more effective than remediation. But the intervention constitutes prevention insofar as it aims to improve future employment prospects before youth leave school and can accrue periods of unemployment.

¹⁵ Since boys offend more than girls (violent-crime arrest control means are 0.13 for boys versus 0.07 for girls), the proportional decrease in arrests is actually larger for girls (50 versus 34 percent, calculated by summing the two coefficients for boys and dividing by the relevant control mean), though the difference is imprecise. Poisson regressions that estimate the proportional effect directly show the same pattern. Take-up rates are nearly identical across genders (74.3 percent for girls and 74.8 percent for boys), so the ITT differences are not driven by differential participation rates.

¹⁶ It is not necessarily the case that this pattern continues linearly out-of-sample; if very high-crime youth are much harder to serve or less responsive than the in-school youth with prior violent-crime arrests in this sample, maximizing violence involvement may not maximize program impacts.

Whether a youth has already been arrested for a violent crime does not change the fact that the program intervenes prior to entry into the full-time labor market. In this respect, both youth with and without a prior violent-crime arrest experience prevention. It may be true that the program could be even more effective if it could identify and serve youth at high risk of violence involvement prior to their first arrest. But the heterogeneous effects suggested by the data cannot speak to that idea, since youth without a prior arrest may just be those at low risk of violence involvement.¹⁷

2.8 Violent Victimization Results

Policymakers care not just about violence that youth commit, but also about the violence committed against youth. It is well established that the risk factors for committing violent crimes are the same risk factors for being victimized (89, 90), and that one's own behavior can be an important determinant of victimization (91). Because treatment reduces violent offending, it might therefore also reduce victimization.

As discussed in the data section, police victimization records are not without problems, particularly because youth have a choice whether to report an incident. It is difficult to distinguish changes in actual victimization from changes in willingness to report – which treatment could affect either by increasing the number of adults around to notice evidence of violence or by changing youths' attitudes towards reporting. Nonetheless, since the total impact of treatment on violence is the sum of both offending and victimization, I report estimates using the available victimization data. To focus on violence, I only count offenses against a person, excluding other non-violent victimizations such as vandalism, forgery, etc. (which are extremely rare in the data).

Table S6 shows ITT and TOT estimates for both violent victimizations alone as well as the sum of violent-crime arrests and violent victimizations. The left panel shows that among the treatment group, there is a proportionally large but statistically insignificant decrease in violent victimization (ITT = -0.0256 incidents per youth, $p = 0.198$, control mean = 0.128). Although the decline in violent victimizations is not statistically significant on its own, it is a substantively large drop (20 percent relative to the control group) in a very socially costly outcome. Combining this change in victimization with the arrest data, the right panel shows a decline of total violence of -0.0652 incidents for each youth offered treatment ($p = 0.014$, a 30 percent decline).

2.9 Under-Reporting of Crime in Arrest Data

One concern about using arrest data to measure criminal activity is that the variable of interest is necessarily under-reported; not every crime results in an arrest.¹⁸ This kind of measurement error is non-classical (the observed y is not just the true y plus a random, mean zero error term). So the standard intuition that measurement error on the left-hand side leaves coefficients unbiased but

¹⁷ The level of criminal involvement among study youth (29 percent of males and 15 percent of females had been arrested at baseline) is about in the middle of prior research on youth employment. The National Supported Work demonstration served very criminally-involved youth (54 percent had been arrested prior to the program) (51). In Job Corps, 33 percent of males had been arrested at baseline and 16.5 percent of females (87), similar to OSP. JOBSTART youth were less than half as criminally involved, with only 12 percent of males and 3 percent of females having baseline arrests (16). And in JTPA, about 20 percent of male youth had a pre-program arrest (88).

¹⁸ The fact that some youth are arrested for crimes they did not commit works in the opposite direction, but is likely to be of much smaller magnitude than the amount of crime that occurs without an arrest.

increases standard errors does not hold. Although the effects of under-reporting will vary depending on the exact form the measurement error takes, this section explores the implications for the simple case where we only observe some constant fraction of the number of true crimes in the arrest data.

Let Y equal the actual number of crimes committed, and suppose the true model is:

$$Y = X\beta + \varepsilon$$

where $E(\varepsilon|X) = 0$. The ordinary least squares estimate of β is:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

and letting $\Omega = \text{Var}(\varepsilon)$, the variance is:

$$\begin{aligned} \text{Var}(\hat{\beta}|X) &= \text{Var}[(X'X)^{-1}X'Y] \\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\ &= (X'X)^{-1}X'\Omega X(X'X)^{-1} \end{aligned}$$

Now suppose that we do not observe the true Y , but rather $Y^* = (1-\delta)Y$, where $0 < \delta < 1$ (so we observe some fraction of the true Y). Then:

$$\begin{aligned} \hat{\beta}^* &= (X'X)^{-1}X'Y^* \\ &= (X'X)^{-1}X'(1-\delta)Y \\ &= (1-\delta)\hat{\beta} \end{aligned}$$

In other words, our estimate using the under-reported dependent variable is too small by the proportional amount of the under-reporting.¹⁹ As a rough ballpark of how much this is likely to affect the main violence results, consider that on average, about half of violent victimizations are reported to the police, and about half of violent crimes reported to the police result in an arrest (43, 44). This would imply that δ is around 0.75 (we observe 1 violent-crime arrest for every 4 violent crimes that actually occur). Since the main point estimate implies a reduction of 4 violent crimes per 100 youth, the true reduction in violent crimes would be $4/(1-0.75) = 16$ per 100 youth.

Importantly, while this would result in a dramatic increase in the program's estimated social benefits, it would change neither our estimate of the proportional effect (a 43 percent reduction), nor our statistical inference. This is because the control mean we observe would also be understated by the same proportional amount as the coefficient, so the estimated percent change in violence is $(\hat{\beta}^*/\text{control mean}^*) = (\hat{\beta}/\text{control mean})$; it is the same regardless of the under-reporting. The same argument is true for the standard error, which is also adjusted by the same factor. To see this, note that:

$$\begin{aligned} \text{Var}(\hat{\beta}^*|X) &= \text{Var}[(X'X)^{-1}X'Y^*] \\ &= \text{Var}[(X'X)^{-1}X'((1-\delta)Y|X)] \end{aligned}$$

¹⁹ This is a common result; see, for example, (92).

$$= (1-\delta)^2 \text{Var}(\hat{\beta} | X)$$

And since the t-statistic for the test of the null hypothesis that $\beta = 0$ is $t(\hat{\beta}) = \frac{\hat{\beta}}{\sqrt{\text{Var}(\hat{\beta} | X)}}$,

$$t(\hat{\beta}^*) = \frac{\hat{\beta}^*}{\sqrt{\text{Var}(\hat{\beta}^* | X)}} = \frac{(1-\delta)\hat{\beta}}{\sqrt{(1-\delta)^2 \text{Var}(\hat{\beta} | X)}} = t(\hat{\beta})$$

In other words, our conclusions about the statistical significance of the coefficient are unchanged, even with under-reported outcomes. So if we thought $\delta = 0.75$ were a good estimate for this setting, we would conclude that the program reduced violent crime by 16 incidents per 100 youth, a statistically significant 43 percent reduction.

This reasoning is also relevant for thinking about the differences across crime types, since under-reporting varies by crime type. Nationally, violent crimes are 2.5 times more likely to result in an arrest than property crimes (43), and as noted in the main text, the overall rate of property-crime arrests seems too low given that property crimes are more common than violent crimes among youth. So one might wonder if the larger amount of under-reporting in property crimes contributes to the fact that we do not see a significant change in property crimes.

The argument laid out here suggests that the point estimate on property crimes is likely to be biased towards zero. However, if we could instead measure the “true” amount of property crime, we would estimate the same proportional change (a 32 percent increase) that would continue to be statistically indistinguishable from zero. In other words, if the measurement error takes the form proposed here – a constant proportional understatement of the true amount of crime – our failure to find a program impact on non-violent crime is not due simply to under-reporting.

2.10 Schooling Results

Table S7 shows ITT and TOT results for several schooling outcomes. All analyses are for youth who were in 8th – 11th grades prior to the program only, since 12th graders’ post-summer schooling outcomes do not exist after graduation. (Not all 12th graders graduate, but since graduation occurs several weeks post-lottery, stratifying on graduation could theoretically undermine the integrity of random assignment.)

The first column of the top panel, where the dependent variable is an indicator for having any attendance during the following school year, shows that there is no difference in re-enrollment rates. Given that 97 percent of the control group returned to school, however, there is not much room for improvement. One might wonder if the high re-enrollment rate suggests that only highly-engaged youth decided to apply to the program. However, the control group’s average number of days attended (middle panel) shows that they are missing about 8 weeks (40 days) of the 181-day school year. In other words, although the program youth have not entirely dropped out, they are also not consistently attending school.

The only statistically significant result in the table is a 3 percentage point (30 percent) decline in summer school enrollment during the summer of the program. It is perhaps not surprising that when given the choice between summer school and earning a paycheck, treatment group youth more often choose the paycheck. If attending less summer school negatively affected later school performance by decreasing academic achievement or preventing grade promotion, this might be

an adverse effect of the program. However, there are no significant differences in days present or GPA, nor in grade retention (not shown). Testing whether the decline in summer school affects progress toward graduation will require longer-term data.

There is no significant decline in summer school attendance the following summer (which might tentatively imply that treatment youth are not working more daytime jobs the following summer, though convincing estimates await Unemployment Insurance data). There are also no significant changes in days present or GPA, regardless of how missing data are treated.

Although the power of this analysis is somewhat lower than for crime due to the exclusion of 12th graders, the confidence intervals still rule out fairly small effects. For example, the 95 percent confidence intervals for days present range from about -6 to 2 days. Relative to a control mean of about 140, a 6-day decrease would be only a 4 percent change in attendance, and a 2 day increase would be a 1.5 percent change. A study of New York City's summer jobs program with a larger study sample and so higher power found statistically significant improvements consistent with the high end of this range (increases of 1-2 days) (81).

Given the magnitude of these estimates, it seems unlikely that school attendance is the key mechanism driving the violence decline. However, since Jacob and Lefgren (31) find that whether or not school is in session has a causal effect juvenile crime, we may still want to quantify how much plausible changes in school attendance (i.e., changes within the confidence interval) could be responsible for the decreased violence. Jacob and Lefgren find that on school days, property crime decreases by 14 percent but violent crime increases by 28 percent (although they find no change in property crime among black youth, who make up the bulk of the OSP sample). Their estimates may not translate perfectly to this setting, since they use teacher in-service days to estimate the effect, meaning that all youth are in or out of school at the same time. They conclude that the concentration of youth in school is what increases violent crime on school days, so their estimates may overstate the violent-crime effect of attendance when individuals are making their own decisions. Nonetheless, as a simplified exercise, their estimates provide a useful basis for a back-of-the-envelope calculation.

The motivating question is how much we could expect violent crime to drop if treatment reduced attendance by the extreme end of the confidence interval: 6 days. If school days have 28 percent more violent crime than non-school days, based on the control group's average school attendance and number of crimes, attending 6 fewer days of school would result in 0.028 fewer violent crime arrests per 100 treatment youth.²⁰ The actual treatment-control difference in the data is two

²⁰ This calculation is as follows: since the school attendance data is only for one school year, I first estimate how many violent crimes the control group committed during the 300 days of the school year (the year was lengthened due to a teachers' union strike but still contained 181 active school days after a revised calendar). For simplicity, I assume that crime is evenly distributed across days. (This is clearly an incorrect assumption which may over-state the amount of crime in on some days, e.g., winter days have fewer crimes, but over-state it on others, e.g., Mondays have fewer crimes than Fridays). Since there are 16 months * 30.4 days per month in the crime data, there are an average of 9.1 violent-crime arrests per 100 youth/486 days = 0.0187 crimes per 100 youth per day, or 5.613 per 100 youth during the 300-day school year. Table S7 shows that the control group attended an average of 140 days of school. From Jacob and Lefgren, we know that $\text{num_viol_crimes}_{\text{per_school_day}} = 1.28 \text{ num_viol_crimes}_{\text{per_nonschool_day}}$. We also know that that $5.613 = 140(\text{num_viol_crimes}_{\text{per_school_day}}) + 160(\text{num_viol_crimes}_{\text{per_nonschool_day}})$. With two equations and two unknowns, we can calculate the number of crimes that happen on the average school and non-school day. Since the exercise is to calculate what the change in violence would be if school attendance were the only difference between control and treatment groups, we can then use these same school and non-school daily crime rates to calculate the total number of crimes if the youth instead attended 6 fewer days of school (134 school

orders of magnitude larger than this estimate (3.95 fewer arrests per 100 youth). Although this calculation is an over-simplification in several respects, the drastic difference in magnitudes nonetheless implies that changes in school attendance are extremely unlikely to explain the changes in violent behavior among OSP youth.

2.11 Benefit-Cost Comparison

A key question from a policy perspective is whether the benefits to society from the program outweigh the program's costs. It is too early to perform a full benefit cost analysis, partly because other key outcomes like employment have not yet been measured, and partly because it is not yet clear whether violence will continue to decrease in the future. Nonetheless, as a descriptive exercise to demonstrate the magnitude of the program's social impact on crime, this section presents some basic calculations to compare the short-term crime benefits to costs. Although there is a great deal of uncertainty in these estimates, reasonable specifications suggest that the program's benefits may already outweigh the costs.

The cost of administering the program for the City of Chicago was about \$3,000 per participant, which includes an average of just over \$1,400 in wages.²¹ 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 (see e.g., 69, 93). This leaves an administrative program cost of about \$1,600 (although if one wanted to separate the costs and benefits that accrue to the government, participants, and society, the wages would appear as a cost to the government and a benefit to participants).

Valuing the benefits of reduced violence requires assigning a monetary value to each type of crime. I use both offending and victimization in the estimates, since an attack on a study youth carries the same social cost as when a study youth attacks someone else. The social costs of crime involving the loss of life (homicide and suicide) tend to swamp changes in other types of crime, and estimates of those costs vary greatly depending on the method used (94, 95). There is no single approach to this issue, so I follow previous work (34, 45) by starting with the full estimated cost of a death in Miller, Wiersema & Cohen (95) and then assessing the sensitivity of the results to trimming the cost in various ways.

Because there are only four fatal crimes in the data,²² estimates of the program's benefits are extremely sensitive to the value assigned to a death. As shown in the first column of Table S8, using the full estimated value of a life (about \$5.2 million in 2012 dollars) results in an adverse

days and 166 non-school days). The total works out to be 0.028 fewer crimes per 100 youth than with the extra 6 days of attendance.

²¹ Note that this is the budgetary cost to the City of Chicago for funding the program. It may understate the costs from a broader perspective, as it does not include the opportunity cost of city staff who might otherwise have worked elsewhere, time donated by program providers beyond what was directly funded by the program, the deadweight loss involved in raising the tax dollars, etc. The exercise here is simply a comparison between the direct monetary cost of the program and the societal benefits from reduced crime.

²² These crimes, which include being a victim or offender in a homicide as well as suicides, are evenly split between treatment and control groups (2 in each group). Since the treatment group is slightly smaller, the proportion of fatal crimes in the treatment group is slightly higher. Given how rare these crimes are, and how close the rates are, this should not be construed as evidence that treatment increases fatal crimes; even one additional data point could reverse the pattern. The very small difference in the proportion of fatal crimes does, however, explain why the estimated social benefit of the program is so sensitive to the treatment of the cost of these crimes.

point estimate of the social costs of participation ($\beta = \$4,212$, $se = \$14,118$).²³ However, the extreme cost of the four very rare events generates a confidence interval so large (-\$23,458 to \$31,883) as to make the point estimate basically uninformative.

Trimming the fatal-crime cost to less extreme values, as in the next two columns of Table S8, reverses the sign of the estimate (suggesting a social benefit rather than cost), and lies just above traditional levels of statistical significance: Setting the cost of a death to the cost of the next most costly crime suggests a program benefit of -\$1,700 (with the negative indicating social savings, $p = 0.117$). In the cost of crime literature, there is also disagreement about whether individuals' use of illegal drugs accrues a social cost or only a cost to the individual (34); setting drug-arrest costs to zero (not shown) raises this estimate of the program's benefit to -\$1,850 ($p = 0.08$).

The last column of Table S8 sets aside fatal crimes altogether. Although this may not be an accurate reflection of the value society puts on a crime like homicide, it does abstract away from the uncertainty of how to value a costly but very rare outcome. This strategy generates a significant estimated benefit of -\$1,878 ($p = 0.062$), exceeding the \$1,600 program cost.

Note the estimated benefits of the program are likely to be conservative in two ways. First, the analysis only counts the decrease in crimes for which youth have been arrested or incidents they have reported to the police. As discussed above, this can dramatically understate the actual change in number of violent incidents. An alternative would be to use a scaling factor to account for the fact that not every crime is known to the police. The cost-benefit analysis of Perry Preschool, for example, multiplies the treatment effect on violent crimes by between 4 and 14, depending on the severity (96).

Second, the social costs of each crime from Miller, Cohen, and Wiersema only count victim costs. They exclude the cost to the criminal justice system of arresting, trying, and potentially incarcerating the offender. They also exclude all of the social costs of contact with the criminal justice system borne by offenders. As such, the social benefits of OSP estimated here are a low-end estimate of the program's social benefits with respect to crime.

In sum, reasonable specifications imply that program benefits from only one outcome (crime) may already outweigh the administrative costs. The uncertainty involved in placing a value on the loss of life, however, means that the substantive conclusion about whether the program's benefits outweigh the costs is sensitive to how four data points are handled. Importantly, it is still early; if the violence reduction persists, the accrual of additional social benefits may make the four fatal crimes less central to the conclusions. Other outcomes not yet measured, such as employment, may also affect the final conclusion about cost effectiveness.

²³ I assign costs to offenses and victimizations based on cost-of-crime estimates from (34, 95). Note that the figures in Table S8 are estimated with two-stage least squares (TOT), since the idea is to compare the benefits of participation with the cost per participant. I could alternatively use ITT estimates, but then would scale the program cost accordingly as well. I do not discount the value of crime reduction over time here, since the outcomes are almost all measured within a year of the program expenditures.

Table S1.

Program participation among entire treatment group (left panel) and program participants (right panel). Program was 5 days per week for 7 weeks, with an 8th week optional. Less than 100% jobs participation among participants likely due to record-keeping error; see supplementary text section 2.3 for details. Treatment arm difference only statistically significant for SEL participation; difference in days attended is just above standard significance levels ($p = 0.12$).

	Offered Treatment			Conditional on Attending		
	Treatment	Jobs-Only	Jobs + SEL	Treatment	Jobs-Only	Jobs + SEL
Any Participation	0.75	0.73	0.77	1	1	1
In Jobs	0.73	0.72	0.73	0.97	1	0.95
In SEL	0.35	0.06	0.63	0.47	0.09	0.83
Total Days	26.07	25.15	26.99	34.99	34.67	35.28
Completed 7 weeks	0.66	0.67	0.66	0.89	0.92	0.86
N	730	364	366	544	264	280

Table S2.

Treatment-on-the-treated estimates from two-stage least squares regressions including baseline covariates and block fixed effects. CCM indicates control complier means. Outcome data from Chicago Police Department administrative arrest records through 16 months post-randomization. Dependent variable for each panel in bold. Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10

	Number of Violent-Crime Arrests	Number of Property-Crime Arrests
Treatment	-0.0533** (0.0227)	0.0167 (0.0157)
CCM	0.121	0.044
	Number of Drug Arrests	Number of Other Arrests
Treatment	0.0073 (0.0230)	-0.0006 (0.0327)
CCM	0.061	0.157

Table S3.

Poisson regression estimates of intent-to-treat program effect, robust standard errors in parentheses. Outcome data from Chicago Police Department administrative arrest and victimization records through 16 months post-randomization. AME shows average marginal effects. Limited set of baseline covariates (block fixed effects, age, number of each type of crime and victimization) included to ensure convergence. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

	Number of Violent-Crime Arrests	Number of Property-Crime Arrests
Treatment	-0.4588** (0.2146)	0.2772 (0.2538)
AME	-0.0368	0.0132
	Number of Drug Arrests	Number of Other Arrests
Treatment	0.1510 (0.3116)	0.0400 (0.1828)
AME	0.0091	0.0058

Table S4.

Intent-to-treat program effects on non-violent arrests by treatment group. Outcome data from Chicago Police Department administrative arrest records through 16 months post-randomization. Difference row shows the magnitude and standard error of the difference in treatment arms. Bottom row shows control means. Baseline covariates and block fixed effects included in all regressions. Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10

	Number of Property-Crime Arrests	Number of Drug Arrests	Number of Other Arrests
Jobs Only	0.0213 (0.0167)	-0.0072 (0.0214)	-0.0271 (0.0310)
Jobs + SEL	0.0036 (0.0137)	0.0179 (0.0284)	0.0260 (0.0384)
Difference	0.0177 (0.0192)	-0.0252 (0.0362)	-0.0531 (0.0493)
CM	0.039	0.049	0.123

Table S5.

Heterogeneous ITT effects by subgroup. Outcome data from Chicago Police Department administrative arrest records through 16 months post-randomization. Baseline covariates (including main effects of covariate in interaction, except for male which is captured by the blocking variables) and block fixed effects included in all regressions. CMs reflect control means for each subgroup. Total n = 1,634, n of subgroups: male = 624, grade 12 = 249, has prior violent-crime arrest = 159. Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10

	Number of Violent-Crime Arrests
Treatment	-0.0359* (0.0185)
Treat*Male	-0.0092 (0.0359)
CM Female	0.070
CM Male	0.132
Treatment	-0.0354* (0.0185)
Treat*Grade 12	-0.0129 (0.0368)
CM 8-11th	0.095
CM 12th	0.066
Treatment	-0.0289* (0.0151)
Treat*Has Prior Viol Arrest	-0.0977 (0.1022)
CM No Viol Prior	0.068
CM Has Viol Prior	0.321

Table S6.

Treatment effect on violent victimization and total violence. Outcome data from Chicago Police Department administrative arrest and victimization records through 16 months post-randomization. Left panel measures violent victimizations; right panel shows effects on all violent crimes (arrests plus victimizations). Bottom row shows control means for ITT and control complier means for TOT. Baseline covariates and block fixed effects included in all regressions. Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10

	Number of Violent Victimization		Number of Violent Crimes (Arrests + Victimization)	
	ITT	TOT	ITT	TOT
Treatment	-0.0256 (0.0199)	-0.0345 (0.0263)	-0.0652** (0.0265)	-0.0878** (0.0351)
CM/CCM	0.128	0.150	0.219	0.272

Table S7.

Intent-to-treat (ITT) and treatment-on-the-treated (TOT) estimates of program impacts on schooling outcomes for youth in 8th - 11th grade at baseline (n = 1,385). "0s Imputed" assigns zeros to anyone not appearing in school records (i.e., assumes dropout). "Means and 0s Imputed" assigns treatment or control group means to students recorded as transferring to another district and zeros to other missing data. 40 youth are missing from attendance file; 154 are missing from grade file. Coefficients from regressions including baseline covariates and block fixed effects. Heteroskedasticity-robust standard errors in parentheses. CM indicates control means; CCM indicates control complier means. Outcome data from Chicago Public Schools administrative data on the 2012-13 school year and prior/following summers. *** p<0.01, ** p<0.05, * p<0.10

	Any Attendance (0/1)		
	2012-13 School Year	2012 Summer School	2013 Summer School
ITT	-0.0019 (0.0092)	-0.0307** (0.0148)	-0.0117 (0.0145)
CM	0.97	0.10	0.08
TOT	-0.0025 (0.0121)	-0.0415** (0.0195)	-0.0158 (0.0192)
CCM	0.97	0.10	0.09
	Number of Days Present (2012-13 School Year)		
	Non-Missing Only	0s Imputed	Means and 0s Imputed
ITT	-1.7718 (1.6099)	-2.0417 (1.9402)	-1.8610 (1.6818)
CM	143.6	139.3	142.2
TOT	-2.3999 (2.1276)	-2.7614 (2.5615)	-2.5170 (2.2212)
CCM	141.6	137.7	140.7
	GPA (2012-13 School Year, 4-point scale)		
	Non-Missing Only	0s Imputed	Means and 0s Imputed
ITT	-0.0444 (0.0382)	-0.0447 (0.0451)	-0.0423 (0.0436)
CM	2.35	2.09	2.15
TOT	-0.0600 (0.0503)	-0.0604 (0.0596)	-0.0572 (0.0576)
CCM	2.28	2.02	2.07

Table S8.

Treatment-on-the-treated estimates of the social cost of program-induced changes in crime (social benefits are negative). Cost of crime estimates from (95), adjusted to 2012 dollars. First column uses full estimated social cost for loss of life (\$5.2 million in 2012 dollars). Second column trims that amount to twice the next most socially-costly crime (sexual assault). Third column sets the cost equal to the next most costly crime. The final column ignores fatal crimes altogether, since they are so rare in the data (4 cases). All regressions include baseline covariates and block fixed effects. Heteroskedasticity-robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.10.

	Full Cost of Fatal Crime	Trim to Twice Next Most Costly Crime	Trim to Next Most Costly Crime	Ignore Fatal Crimes
Treatment	\$4,212 (14,118)	-\$1,522 (1,295)	-\$1,700 (1,083)	-\$1,878* (1,004)