NBER WORKING PAPER SERIES

PREDICTING AND PREVENTING GUN VIOLENCE: AN EXPERIMENTAL EVALUATION OF READI CHICAGO

Monica P. Bhatt
Sara B. Heller
Max Kapustin
Marianne Bertrand
Christopher Blattman

Working Paper 30852
http://www.nber.org/papers/w30852

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2023, Revised October 2023

The research had support from a wide philanthropic community, including: Arnold Ventures; the Partnership for Safe and Peaceful Communities; JPMorgan Chase; the Chicago Sports Alliance; and the Institute for Firearm Injury Prevention at the University of Michigan. A huge team of research staff made the study possible, with enormous thanks to Damilare Aboaba, Xander Beberman, Brenda Benitez, Ryan Carlino, Ran Cheng, Binta Diop, Brandon Domash, Mara Heneghan, Miguel Hernandez-Pacheco, Megan Kang, Leah Luben, Connor McCormick, Melissa McNeill, Evelyn Morris, Danielle Nemschoff, Michelle Ochoa, Priyal Patil, Mark Saint, Michael Tatone, Diamond Thompson, and Nathan Weil. We are indebted to Heartland Alliance and its extraordinary leadership team, to Zubin Jelveh and Ben Jakubowski for their work developing the prediction model, to Roseanna Ander, and to the remarkable local community organizations that made READI happen: Centers for New Horizons, Cure Violence, Heartland Englewood Outreach, Heartland Human Care Services, the Institute for Nonviolence Chicago, Lawndale Christian Legal Center, North Lawndale Employment Network, and UCAN. We are also grateful to our data providers, the Chicago Police Department, Cook County Sheriff's Office, and the Illinois Department of Corrections; to Betsy Levy Paluck and Andrew V. Papachristos for their insights in READI's early stages; and to Jennifer Doleac, Peter Hull, Lawrence Katz, Doug Miller, and Jens Ludwig for helpful comments. All opinions and any errors are our own and do not necessarily reflect those of our funders, implementing partners, data providers, or other government agencies. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2023 by Monica P. Bhatt, Sara B. Heller, Max Kapustin, Marianne Bertrand, and Christopher Blattman. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Predicting and Preventing Gun Violence: An Experimental Evaluation of READI Chicago
Monica P. Bhatt, Sara B. Heller, Max Kapustin, Marianne Bertrand, and Christopher Blattman
NBER Working Paper No. 30852
January 2023, Revised October 2023
JEL No. C53,C93,I38,J08,K42

ABSTRACT

Gun violence is the most pressing public safety problem in American cities. We report results from a randomized controlled trial (N = 2, 456) of a community-researcher partnership called the Rapid Employment and Development Initiative (READI) Chicago. The program offered an 18-month job alongside cognitive behavioral therapy and other social support. Both algorithmic and human referral methods identified men with strikingly high scope for gun violence reduction: for every 100 people in the control group, there were 11 shooting and homicide victimizations during the 20-month outcome period. Fifty-five percent of the treatment group started programming, comparable to take-up rates in programs for people facing far lower mortality risk. After 20 months, there is no statistically significant change in an index combining three measures of serious violence, the study's primary outcome. Yet there are signs that this program model has promise. One of the three measures, shooting and homicide arrests, declines 65 percent (p = 0.13 after multiple testing adjustment). Because shootings are so costly, READI generates estimated social savings between $182,000 and $916,000 per participant (p = 0.03), implying a benefit-cost ratio between 4:1 and 18:1. Moreover, participants referred by outreach workers—a pre-specified subgroup—show enormous declines in both arrests and victimizations for shootings and homicides (79 and 43 percent, respectively) that remain statistically significant even after multiple testing adjustments. These declines are concentrated among outreach referrals with higher predicted risk, suggesting that human and algorithmic targeting may work better together.

Monica P. Bhatt
University of Chicago
Crime and Education Labs
33 North LaSalle Street
Suite 1600
Chicago, IL 60602
mbhatt@uchicago.edu

Sara B. Heller
University of Michigan
Department of Economics
611 Tappan Street, Lorch Hall Room 238
Ann Arbor, MI 48109
and NBER
sbheller@umich.edu

Max Kapustin
Department of Economics
Brooks School of Public Policy
Cornell University
Martha Van Rensselaer Hall
Ithaca, NY 14853
kapustin@cornell.edu

Marianne Bertrand
Booth School of Business University
of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
marianne.bertrand@chicagobooth.edu

Christopher Blattman
Harris School of Public Policy
The University of Chicago
1307 E 60th St
Chicago, IL 60637
and NBER
blattman@gmail.com

A randomized controlled trials registry entry is available at https://osf.io/ap8fj/

1 Introduction

Over 170 Americans are shot each day, with young Black men dying of gun homicide—by far their leading cause of death—at almost 20 times the rate of their White peers (CDC, 2020).[1] There are strong arguments that more effective policing and punishment can reduce community gun violence (see, e.g., Braga and Cook, 2023). But there is also a concern that common law enforcement strategies—aggressive policing that prioritizes street stops and low-level arrests, for example, or much greater use of prisons (Harcourt, 2005; Raphael and Stoll, 2013)—can impose high social costs, especially on the same communities that already bear the burden of gun violence itself. This concern has fueled demand for ways to reduce shootings without the harms of overly aggressive or poorly targeted law enforcement.[2]

One feature of community gun violence—its concentration—may be key to reducing it. In Chicago, for instance, five neighborhoods accounted for a quarter of the shootings in 2022, despite containing a tenth of the city's population. Within such neighborhoods, moreover, the number of people involved in gun violence is small (e.g., Braga, 2003). If the *ex ante* risk of being a shooting victim or offender is concentrated enough, then intervening with a small set of people could meaningfully and cost-effectively reduce shootings (Abt, 2019; Green et al., 2017; Heller et al., 2022).[3]

Any individually-targeted intervention must overcome two challenges. The first is identifying and engaging people at high risk of gun violence. This a difficult prediction problem given the complex determinants of shootings (Chandler et al., 2011; Heller et al., 2022; Wheeler et al., 2019), and a difficult practical problem given the extreme levels of trauma and disconnection in this population (Fagan and Wilkinson, 1998; Anderson, 1999).

[1] Rates are for non-Hispanic Black and White men ages 18-34. These numbers exclude suicides, accidents, and shootings by police.

[2] See, e.g., Ang (2021); Geller et al. (2014); Jones (2014); Chalfin et al. (2022); Pattillo et al. (2004).

[3] A complement to this targeted, individual approach is to address the "root causes" of gun violence, such as concentrated disadvantage and access to guns. A targeted approach can be implemented and help quickly, while making structural changes will take more time (see, e.g., Haveman et al. (2015) and *New York State Rifle & Pistol Association, Inc. v. Bruen*, 597 U.S. (2022)).

The second challenge is finding effective ways to reduce the risk of being involved in a shooting. Many cities are now funding social services that try to overcome both challenges.[4] Unfortunately, we know relatively little about how to find, let alone engage, those at very high risk of gun violence without law enforcement involvement. Nor is there much rigorous evidence on what social services actually reduce this risk (see Section 2.1 and Appendix A.1).

This paper evaluates a community-researcher partnership designed to tackle both challenges: the Rapid Employment and Development Initiative (READI) Chicago. READI operated in five of Chicago's highest-violence neighborhoods. The program sought to identify men at the very highest risk of shooting involvement using three referral pathways: (1) a machine learning algorithm based on administrative arrest and victimization records; (2) referrals from local outreach workers; and (3) screening among those leaving prison and jail. Over three years, 2,456 men were randomly assigned either to a READI offer or to a control group free to pursue other services.

Men assigned to READI were offered 18 months of subsidized, supported work combined with group cognitive behavioral therapy (CBT). The job was designed to provide several elements: a stable source of income to deter illegal work, an incentive to participate in the therapy, a place to build and reinforce new skills and norms, and a reason to spend less time in dangerous settings. Meanwhile, the CBT-informed programming was designed to foster several complementary behavior changes: to help participants reflect on their own thinking, practice less harmful responses in dangerous situations, and promote more positive behaviors and identities. Due to the significant barriers to participation this population faces, READI also provided participants with referrals to housing, substance abuse, mental health, and legal services when needed.

With respect to the first challenge—identifying and engaging people at very high risk of shooting involvement—READI was a clear success. Prior to program referral, 35 percent of

[4] See, e.g., https://www.chicago.gov/content/dam/city/sites/public-safety-and-violenc-reduction/pdfs/OurCityOurSafety.pdf, https://www.phila.gov/2021-04-14-how-the-city-is-addressing-gun-violence-2021-update-to-the-roadmap-to-safer-communities/, https://www.oaklandca.gov/topics/oaklands-ceasefire-strategy, and https://monse.baltimorecity.gov/sites/default/files/MayorBMS_Draft_ViolenceReductionFrameworkPlan.pdf.

men in the study had been shot and 98 percent had been arrested, with an average of over 17 prior arrests. Staggeringly, in the 20 months after randomization, there were 11 shooting and homicide victimizations for every 100 men in the control group—52 times more than among average Chicagoans, and 2.7 times more than among other similarly-aged men in the five READI neighborhoods.

Despite many barriers to participating, 55 percent of men assigned to treatment attended at least one day of programming. Overall, participants worked an average of 30 percent of the hours available, prior to the suspension of in-person programming in March 2020.[5] The subset of participants who continued to work remained engaged, working 75 percent of the weeks available to them while in-person work was occurring. This rate is comparable to interventions for much lower risk populations (such as high school boys) and for much shorter transitional job programs (Heller et al., 2017; Redcross et al., 2016).

On the second challenge—reducing serious violence—the results provide reason for both caution and optimism. We track three measures of serious violence involvement over a 20-month outcome period using matched administrative data: (1) shooting and homicide victimizations; (2) shooting and homicide arrests; and (3) other serious violent-crime arrests, such as robbery and aggravated battery.[6] Our primary pre-specified outcome is a standardized index that averages all three measures of serious violence with equal weights. We also specified several secondary analyses, including how READI affects the index components, as well as an index of all crime and violence incidents weighted by their social costs.

There is no detectable impact of READI on the primary outcome—the simple average of the three serious violence measures. The estimated effect of treatment on the treated (TOT) is -0.049 standard deviations ($p = 0.26$). When we break the index into its three components, however, we find suggestive evidence that READI reduced arrests for shootings

[5] The READI study was in the field from August 2017 through October 2021. At the start of the COVID-19 pandemic, CBT sessions shifted online and in-person work was temporarily suspended, though some payments continued; see Section 2.3. Because our outcome window is 20 months, about 76 percent of person-day post-randomization observations occurred before the pandemic.

[6] Our pre-analysis plan is available at `https://osf.io/ap8fj/`.

and homicides. Relative to men in the control group who would have started programming if offered (control compliers), READI participants had 65 percent fewer shooting and homicide arrests (2.2 fewer per 100 participants). This result is statistically significant on its own, but not after adjusting inference for the three hypothesis tests involved in breaking the index into its components (unadjusted $p = 0.05$, adjusted $p = 0.13$). Results for the other two components are less precise. Point estimates show that participants had 12 percent fewer shooting and homicide victimizations but 11 percent more arrests for other types of serious violent crime. Confidence intervals, however, are too wide to draw clear conclusions (adjusted $p = 0.8$ and $p = 0.7$, respectively).

When we weight incidents of crime and violence by the costs they impose on society, however, we estimate that READI reduced these social harms by at least $182,000, and perhaps by as much as $916,000, per participant—about a 50 percent decline ($p = 0.03$). The increased precision comes from the fact that the large decreases in violence are concentrated in the most socially costly outcomes. Using a range of assumptions, these estimates imply READI's benefit-cost ratio is at least 4 to 1, and could be as high as 18 to 1.

READI also generated heterogeneous treatment effects. We pre-specified that we would analyze results by referral pathway, and impacts differ significantly from each other ($p = 0.03$). Participants referred by outreach workers saw serious violence involvement fall by 0.13 standard deviations (adjusted $p = 0.03$), driven by large and statistically significant reductions in shooting and homicide arrests (79 percent, adjusted $p = 0.03$) and victimizations (43 percent, adjusted $p = 0.08$) relative to control compliers in the same pathway. Hence, READI was more effective at reducing serious violence among outreach referrals.

It is harder to say *why* impacts were larger for this pathway. The results are consistent with those men receiving a higher dose of programming, but also with them being more responsive to it. Outreach workers were asked to refer men at highest ex ante risk of gun violence involvement, which we refer to as selection on $\hat{Y}(0)$, or the predicted level of Y in the absence of treatment. Interviews suggest that, as anticipated, outreach workers also considered referrals' expected gains from treatment, which we refer to as selection on $\beta = Y(1) - Y(0)$, or the treatment effect of READI. Program staff frequently reported filtering out men who they felt were not "ready"—not open to a change of lifestyle or facing too many barriers to participation, and thus unlikely to engage in the program.

In exploratory analysis, we unpack treatment heterogeneity between outreach and algorithm referrals. Men referred by the algorithm had, on average, a higher predicted risk of future gun violence involvement. Yet at almost all levels of predicted risk, outreach referrals were subsequently involved in gun violence at higher rates. This suggests that outreach workers' screening methods successfully incorporated risk factors for $Y(0)$ that the algorithm could not observe. However, outreach workers were only partially successful in identifying men with a high β; violence declined only among the subset of outreach referrals who also had the highest predicted risk. Together, these patterns suggest that outreach workers were not simply selecting on expected gains. Rather, the combination of higher observable risk *and* outreach-identified unobservables appears to predict treatment responsiveness. In other words, human and algorithmic referral mechanisms worked better together than alone.

From the perspective of scientific hypothesis testing, the mixed program impacts we document make it difficult to give a definitive answer about how READI changed behavior. It likely reduced shooting and homicide offending (as measured by arrests) overall, as well as drastically lowered shooting and homicide victimization for men referred by outreach. But we fail to reject the null for all forms of serious violence across all subgroups. Future research replicating and refining READI's approach would be valuable to learn whether a combination of work and CBT can reduce gun violence among the men at highest risk of it.

From a policy perspective, however, binary hypothesis tests may not be the most useful basis for decision-making. As others have argued, policymakers should weigh the importance of the outcome, the uncertainty of the estimates, and the availability of other interventions and evidence (Imbens, 2021; Manski, 2019; Ziliak and McCloskey, 2008). From this perspective, a few aspects of the READI results are worth highlighting. For the primary index of serious violence, 74 percent of the treatment effect's confidence interval is below zero. Policymakers can weigh this against their level of uncertainty about the effectiveness of other social service approaches to reduce gun violence, as well as the increased social costs that aggressive law enforcement responses can generate. They can also use our estimate of READI's benefit-cost ratio—in essence, an importance-weighted sufficient statistic, as in Viviano et al. (2021)—which suggests that society values READI's impact on violence between 4 and 18 times more than the cost of running the program.

One clear lesson from these results is the potential for a targeted intervention to affect the total amount of gun violence in a city. Despite being less than 0.01 percent of Chicago's population, the 2,456 men in the study sample would have contained about 6 percent of Chicago's shooting and homicide victims during an average 20-month period in the absence of READI, costing society between $711 million and $3.6 billion.[7] And despite being disconnected from and distrustful of many social institutions, these men proved willing to engage in READI. The fact that it is possible to identify and engage a relatively small group at such elevated risk of socially costly outcomes emphasizes the potential of continuing to experiment with approaches to help this extremely disconnected and under-served population.

2 Experimental sample and intervention

2.1 Context & research questions

READI was designed in response to an unprecedented 60 percent spike in Chicago's homicide rate from 2015 to 2016 (Kapustin et al., 2017).[8] As in many cities, Chicago's shootings are extremely concentrated in neighborhoods with many low-income residents and historically high rates of violence. Within such neighborhoods, gun violence appears to be further

[7] We calculate the 6 percent figure using the number of shooting and homicide victimizations during the 20-month outcome period in the control group (140) as the counterfactual for the number in the treatment group without READI, compared to the number in Chicago during an average 20-month period between August 2017 and August 2021 (about 4,600).

[8] From 2016–21, Chicago experienced an average of 23 homicides per 100,000 people annually, the vast majority due to guns. This rate is more than 8 times higher than those in Los Angeles or New York, comparable to those in Philadelphia and Milwaukee, and around half of those in cities such as St. Louis or Baltimore.

concentrated among a small group of people; a wide range of evidence suggests that only a tiny percentage of the population engage in serious violent crime (Green et al., 2017; Braga, 2003; Farrington et al., 2006; Wolfgang and Tracy, 1982; Abt, 2019).

Using targeted interventions to tackle concentrated gun violence is a longstanding idea, both in policing and among community violence interventions, or CVIs (Sherman and Rogan, 1995; Braga et al., 2001, 2018, 2014; Skogan et al., 2008; Butts et al., 2015a). Appendix A.1 discusses the research about targeted policing as well as CVI approaches. One of the most common, community-wide violence interruption, involves intervening in and mediating active disputes. Such mediation is qualitatively and theoretically important to reducing violence, but it is also challenging to evaluate due to the difficulty of finding counterfactual communities (Farrell et al., 2016; Roman et al., 2018). A review of the evidence characterizes it as mixed (Butts et al., 2015a). A newer and complementary set of CVIs proactively offer preventative services to specific people or groups at high risk of gun violence involvement. This kind of targeted, tertiary prevention also has qualitative and theoretical promise, but there is no causal evidence so far of its effectiveness.

Separately, causal evidence exists on READI's two main program components—supported work and CBT—albeit for purposes and populations quite different than CVIs meant to reduce shootings. Studies of transitional jobs programs suggest that they are unlikely to reduce crime and violence on their own, but that strategies combining jobs and enhanced services such as CBT are more effective (Cummings and Bloom, 2020; MDRC, 2013; Redcross et al., 2016). And while CBT-informed programming alone can reduce violence involvement, there is some evidence it may be more effective when paired with an economic intervention (Heller et al., 2017; Blattman et al., 2022, 2017; Lipsey et al., 2007; Wilson et al., 2005; Dinarte and Egaña del Sol, 2019; Arbour, 2022). While both of READI's core program elements have shown promise, neither has been evaluated on a population as disconnected and at as high risk of gun violence as the one that READI aimed to serve.

Any CVI that seeks to prevent shootings by intervening with specific people must meet two criteria to be successful: (1) identify and engage a group of people at high enough *ex ante* risk of gun violence for it to be feasible to reduce shootings among them, and (2) reduce their risk of being involved in a shooting. Both criteria pose significant challenges.

Identifying this population is a difficult prediction problem (Berk et al., 2009; Chandler et al., 2011; Heller et al., 2022; Wheeler et al., 2019). Existing research provides little guidance about whether and which observable and unobservable characteristics can identify specific people at very high ex ante risk, particularly given the idiosyncrasies inherent in human behavior and in the likelihood of being involved in gun violence itself (i.e., whether a person is hit or missed). To the extent that gun violence involvement risk is transitory, this population may also change over time.

Once identified, finding these individuals can be extraordinarily challenging, especially if they keep a low profile to avoid encounters with police or their opposition. Once found, they may be disconnected from—and skeptical of—societal institutions and offers of help, in addition to facing many logistical barriers to participation such as housing instability, substance use disorders, and safety concerns about exposing themselves to certain people or places (Fagan and Wilkinson, 1998; Anderson, 1999).

Most CVIs use outreach workers' local relationships and expertise to find and engage clients—usually young men with high rates of recent violence exposure. But without a random comparison group, evaluations cannot tell us clients' gun violence risk in the absence of services. It is also unclear how many people at high gun violence risk are missed by relying solely on expert referrals, or whether those referrals have the highest gains from participation. Beyond the challenges of identifying, finding, and engaging the relevant population, we lack evidence on what kinds of interventions can reduce shooting involvement.

As a result, the READI study was not designed with a sole focus on impact evaluation. Rather, we set out to answer three questions: (1) Can we identify men at high enough risk of future gun violence that there is scope to reduce shootings?; (2) Will they participate in a pro-social intervention?; and (3) Will a combination of supported work and CBT reduce their involvement in serious violence?

2.2 Sample selection, referral pathways, and randomization

READI was a partnership between several organizations: Heartland Alliance, an anti-poverty and human rights non-profit based in Chicago which designed, developed, and managed READI; four organizations specialized in outreach; three organizations specialized in employment and CBT-based programming; the principal investigators; and the University of Chicago Urban Labs.[9] All parts of the program were developed collaboratively with input from community workers and a Participant Advisory Council.

Eligibility READI aimed to recruit men 18 and over at the highest risk of gun violence involvement. It focused on five of 77 Chicago neighborhoods with the highest levels and rates of gun violence (Figure I). Providers grouped these neighborhoods into three sites: Austin/West Garfield Park, North Lawndale, and Englewood/West Englewood.[10]

Given READI's explicit goal of serving men at the highest risk of gun violence, it is worth emphasizing that being at high ex ante risk for an outcome does not necessarily mean being responsive to an intervention designed to reduce it—that is, having a high $\hat{Y}(0)$, the predicted level of gun violence involvement absent treatment, is not necessarily the same as having a high β, the treatment effect of READI. Understanding the relationship between $\hat{Y}(0)$ and β is nonetheless important for future interventions. To encourage variation in participants that would allow us to study this, we designed three pathways to recruit men on a rolling basis, described below, with additional details in Appendix A.3.

Referral pathways The *algorithm pathway* used administrative police data to predict the risk of being involved in gun violence as a victim or an arrestee over the next 18 months (the

[9] Program implementation started in partnership with Centers for New Horizons, Cure Violence, Heartland Human Care Services, the Institute for Nonviolence Chicago, Lawndale Christian Legal Center (LCLC), North Lawndale Employment Network, and UCAN. In September 2018, Heartland Alliance took over outreach services from Cure Violence in Englewood. And in April 2021, UCAN took over outreach services from LCLC in North Lawndale.

[10] READI continues in these sites using a modified program model for non-study participants.

"risk score").[11] Each time program slots became available for algorithm referrals, men with the highest risk scores who met READI's eligibility criteria were referred for randomization. This algorithmic approach is useful insofar as observables can successfully predict a person's involvement in gun violence ($\hat{Y}(0)$), but it will miss risk driven by unobservables or fast-moving situations that are not reflected in the police data used in the algorithm.

To capture some of these unobservables and allow for possible selection on treatment responsiveness (β), the *outreach pathway* sourced referrals from outreach workers with extensive on-the-ground experience in the READI neighborhoods. These workers are privy to local information that may be absent from police records, but may be limited by the scope of their social networks or their incentives to offer particular people services and fill caseloads. They were asked to refer men at the highest risk of gun violence.

Finally, the *re-entry pathway* identified men leaving jail or on parole who may be missed by both outreach workers and the algorithm, and who may be at a particularly sensitive transition point. The re-entry pathway took the longest to become operational due to the complicated logistics involved with operating within carceral facilities. Because COVID-19 ended study recruitment early, this pathway is considerably smaller than the other two and than we initially intended. As such, we focus on differences between the first two pathways but report re-entry results separately for completeness.

Having three referral pathways allows us to assess how the different ways of identifying men at the highest risk of shooting and being shot performed: whether observables are enough to predict future gun violence with a machine learning algorithm, whether the algorithm identified a different set of people than the outreach workers, and why those groups might differ (i.e., whether on-the-ground knowledge could capture unobservables in a way that improved program targeting, and whether human decision-makers chose to select not only on gun violence risk but also expected responsiveness to the program).

[11] We predict risk scores for those with sufficient recent police contact; see Appendix A.3.1. Heller et al. (2022) also provides a full description and analysis of a related prediction model, incorporating lessons from READI's prediction model.

Randomization READI solicited referrals on a rolling basis from August 2017 to March 2020. New referrals ended early at the start of the pandemic, shifting the sample size to 2,456 from the original target of 3,000. Rolling referrals were made from all three pathways as program slots became available, both to accommodate READI's growing capacity to absorb new participants over time, and to focus on the people at highest risk, who may change over time. Outreach referrals began in August 2017, algorithm referrals in December 2017, and re-entry referrals in August 2018.

The randomization process varied slightly by pathway but followed the same general structure (see Appendix A.3). In all cases, we randomized at the individual level with a treatment probability of one half, within strata defined by site, referral pathway, and randomization date.[12] After receiving referrals via the outreach or re-entry pathways and matching them to administrative data, or after identifying men with the highest risk scores via the algorithm pathway, we screened out anyone who had previously been randomized, was incarcerated, or who had died since their referral.

2.3 The READI program

READI is a bundled intervention designed to disrupt four proximate causes of gun violence. First is the instrumental use of violence in illegal markets, where illicit organizations lack legal ways to enforce contracts or compete for market share. Second is rational reputation building, where people use violence to signal strength as a way to deter future attacks in a dangerous environment. Third is reciprocity, where violence is a means to punish real or perceived slights, especially in settings where the legal system is not viewed as legitimate or just. And fourth is "irrational" behavior arising from mistakes and misperceptions—instances where fast decision-making combined with fear, anger, or persistently biased beliefs about others' intentions can result in violent escalation.[13]

READI's key components were designed to address these proximate causes of gun vi-

[12] In practice there is slight variation in treatment probability within strata; see Appendix A.3.4.

[13] For an extended theoretical discussion, see Blattman (2022) and Abt (2019). For a historical analysis of how these dynamics have played out in Chicago, see Aspholm (2020).

olence. Note, however, that because it was designed and implemented extremely quickly to respond to crisis-level violence in Chicago, READI's development was a learning process with the model changing slightly over time.

Initial outreach Engagement in all pathways began with outreach workers trying to locate men assigned a READI offer and persuading them to participate. These men were usually mistrustful of organized programming. Once located, outreach workers tried to convince these men to join READI, including by helping them obtain documentation to work legally or negotiate a truce with members of opposition groups in the program. This process of building trust and readiness could take days or months, depending on the person and their existing relationships with the outreach organization. Once a person was willing and ready to begin, the outreach worker connected them to their local READI employment organization. Attending an orientation seminar and signing job paperwork denotes the beginning of formal participation in READI—what we define as "taking up" the program.

Supported, subsidized work To incentivize participation and provide an alternative to work in illegal, violence-prone markets, READI's first component was supported, subsidized work. Participants could earn money at worksites 5 days per week (29.5 hours total), for up to 18 months.[14] READI was explicitly *not* intended as a transitional job program to rapidly place men into full-time work. It was focused on violence reduction. Nonetheless, READI was informed by best practices, including a "career pathway" approach with four stages based on a participant's progress (Appendix Figure A.I).

During the first stage, participants were typically assigned to crews performing outdoor work (such as park cleanup) or other basic services (such as packing meals for food pantries). They received transportation to and from the worksite, as safety was a core challenge. Later stages offered participants a greater variety of jobs and more independence, potentially including subsidized placements with local employers (e.g., some participants

[14] READI recognized that its target population faced many barriers to employment, so men who stopped attending work were allowed to later resume participating; see Appendix A.4.1.

eventually worked in a local vehicle seat factory).

This tiered structure allowed participants to expand their skill set and earning potential over time, which prior research suggests is important for keeping them engaged. Participants in the first stage received a minimum hourly wage, initially $11 but rising during the course of the study to match changes in the local minimum wage. Advancing to each stage was accompanied by a wage increase, among other benefits (Appendix Figure A.II).

Given the mixed results of prior jobs programs, even with populations at much lower risk of violence (Cummings and Bloom, 2020; MDRC, 2013; Redcross et al., 2016), the job component was not expected to reduce violence on its own. Nonetheless, over dozens of interviews and focus groups, participants and program staff said the job was a crucial incentive to participate (see Section 3.3 and Appendix A.8.1 for qualitative data and methods). The job was also designed to complement therapy by providing a place to practice and reinforce new thinking and behaviors, guided by staff at the worksites. Finally, given the level of violence involvement the population was expected to have, simply keeping participants busy during the week could potentially have a substantial incapacitation effect.

CBT-informed curriculum To reduce the kind of "hot" decision-making that leads altercations to escalate, and to help build new social identities and norms around violence, social interactions, and illegal employment, the second element of READI involved cognitive behavioral therapy (CBT). CBT is an approach for reducing maladaptive beliefs and behaviors, and for promoting positive ones. Its methods can be applied to a range of behaviors, and CBT-informed therapies have been successful at reducing symptoms of depression, anxiety, phobias, traumatic stress, and hostility (Beck, 1979; Beck and Dozois, 2011).

CBT-informed therapies like READI teach participants that thoughts can influence actions, and help them practice actions designed to shape those thoughts. In 90-minute sessions for three mornings each week, program staff facilitated conversations and conducted exercises with small groups of men designed to help them become more conscious of their automatic thoughts, particularly inaccurate or negative ones about themselves or others that could lead to violence or other behaviors that are inappropriate in a given situation. Facilitators taught participants techniques to recognize these thoughts and respond to them in ways that are more constructive and less harmful. Both in group sessions and outside of them as "homework" (including at worksites), men had opportunities to practice these techniques on tasks of increasing difficulty. Through such "learning by doing," CBT can gradually modify participants' behavior and thinking.

Participants also received individual and small-group professional development and information sessions the remaining two mornings per week. They received a $25 gift card for each CBT and personal development session attended. Afterward, they departed to their worksites for the day. The CBT sessions were available for all 18 months of READI and were a requirement for participation in supported work.

Ongoing outreach support & referrals to other services Initially, READI did not plan for outreach staff to be continually engaged with participants, but rather to focus on recruitment. Within the first weeks, however, outreach workers were finding it necessary to devote a significant share of their time to this engagement, and so it was formally incorporated into the program (see the qualitative analysis in Appendix A.8.5 for details on how and why outreach workers engaged with participants). Throughout our qualitative interviews and observations, READI staff and participants both emphasized the extreme struggles participants faced in continuing with the program day-to-day. These include frequent fights, confrontations with program staff, and other serious disturbances on site—including occasional incidents of gun violence from outsiders and participants. Outreach workers provided almost continuous support, applying their training in conflict mediation, de-escalation, and restorative justice, including at worksites.

In addition to these safety issues, participants faced other serious challenges including episodic homelessness, family quarrels, financial difficulties, arrests and other legal troubles, parole commitments, physical and mental health struggles, and other issues that hindered participation. Outreach and program staff helped participants navigate these problems, including by making referrals to organizations providing substance abuse treatment, housing, or legal services.[15] Often, these situations served as opportunities to employ skills acquired in the CBT sessions and entrench the new behaviors as habits.

Services available to the control group Men in the study were free to access alternative programming. It is reasonable to assume that some did, particularly men in the control group who lacked access to READI. To our knowledge, no other provider offered programming of READI's length or intensity, with a similar combination of services, delivered to men facing a similarly high level of risk during the study period and in the READI neighborhoods.[16] Though the number of organizations in Chicago offering individual programming intended to reduce gun violence has grown since READI started in August 2017, few provide jobs, let alone for 18 months or in combination with CBT.

Unfortunately, due to their level of risk, mobility, and distrust, it is impossible to survey control group men to learn about their program participation directly. Service providers also tend to be protective of client identities. We can, however, indirectly learn about this from the control group's primary conduit to alternative programming: outreach workers. Members of the research team conducted focus groups with nearly all program staff, as well as 220 hours of field visits where explicit attention was paid to engagement of the control group. Appendix A.8 summarizes the analysis. Briefly, several staff described an intense effort to assist the highest-risk men, largely because they felt these men were in life-or-death situations. These outreach staff tried to stay in touch with control group men, mentor them, and connect them to temporary work agencies or mentorship programs. While there are scattered reports of success at these efforts, one outreach worker expressed a common view that "there is nothing like READI around here." Staff efforts to connect control group men

[15] Criminal legal services were focused on dealing with a client's prosecution. As such, they were unlikely to affect our measure of violent offending, which occurs prior to prosecution, at the point of arrest.

[16] The closest comparable program is Chicago CRED, which provides jobs, life coaching, trauma counseling, and education to a similar population of young men but operates mostly in the Roseland and West Pullman neighborhoods on Chicago's South Side: `https://www.ipr.northwestern.edu/documents/reports/ipr-n3-rapid-research-reports-cred-outreach-jan-22-2021.pdf`

with other programs were also heterogeneous, with some staff reporting that they did not have the bandwidth to provide continued support.

Any support provided to control group men, however modest, would likely lead us to underestimate READI's treatment effects and overstate its costs. This is probably most acute for men referred by outreach workers, as their pre-existing relationships could have facilitated ongoing contact and access to alternative services. In contrast, outreach workers typically had no information about control group men from the algorithm pathway.

Impacts of the COVID-19 pandemic on program delivery In March 2020, all randomization, qualitative data collection, and in-person READI programming was suspended. The CBT sessions transitioned to being held online. Supported work was paused altogether, with participants receiving "standby pay" from March through July 2020. In August 2020, in-person work and CBT sessions resumed. Participants had the option of going to work in person or earning up to $50 per day attending professional development sessions remotely.

While it did not undermine the study's internal validity, the pandemic affected the study in two ways. First, our sample size is almost 20 percent smaller than initially intended due to randomization ending early. Second, services were severely disrupted for those already participating, and the resumption of in-person work after a pause of many months made it difficult to re-engage some men. Since the meaning of "participation" changed with the pandemic, in the main text we report overall take-up rates, program hours, and earnings, but limit reports of retention to the pre-COVID period. Appendix A.5.3 provides more detail.

3 Data and empirical strategy

3.1 Outcomes and data

The inherently unobservable or "latent" outcome of interest is a person's involvement in serious violence, either as a perpetrator or a victim. We proxy for both kinds of violence involvement using administrative arrest and victimization records from the Chicago Police Department (CPD). The main advantage of police data is that we can use them to follow a large number of study members over a long period of time, with minimal sample attrition and relatively low cost. This makes the study feasible to conduct, as it would not be possible to track behavior via surveys over long periods of time with such a disconnected and difficult-to-locate population. Administrative records also avoid social desirability bias that might prevent people from honestly disclosing their serious violence involvement.

Police data also have serious limitations. One is that victimizations only appear when victims are willing to report incidents to the police. That said, because all healthcare workers in Illinois are legally required to report shooting victimizations to local police (20 ILCS 2630/3.2), and because shooting victims are very likely to seek medical care, under-reporting of non-fatal shooting victimizations is likely to be minimal.[17] Similarly, homicide victimization is widely thought to be mostly free of under-reporting in police data (Loftin and McDowall, 2010; Carr and Doleac, 2016).

A more serious challenge is the use of arrests. Arrests likely understate offending due to low clearance rates, even for serious crimes: only 26 percent of homicides and 5 percent of non-fatal shootings in Chicago in 2016 resulted in an arrest (Kapustin et al., 2017). Arrests are also subject to potentially biased police decisions and may be mistaken or wrongful. The chance of an offense resulting in an arrest (or the chance of a false arrest) may also vary by demographics or neighborhood. While these issues introduce error into arrest measures that could make it harder to detect treatment effects on offending, they affect both treatment and control groups, and so should not bias impact estimates. The key assumption is that treatment does not change the probability of arrest conditional on actual criminal behavior.[18]

Finally, the data in this paper do not capture victimizations or arrests outside of Chicago. However, if READI's steady paycheck increased the time participants spent in Chicago, then their violence involvement would be more likely to appear in our data, causing us to

[17] Some under-reporting may still occur if shooting victims self-treat or seek care from medical providers outside of Chicago who may not report such victimizations to the CPD. However, based on our conversations with violence prevention, medical, and law enforcement practitioners in Chicago, we think the magnitude of such under-reporting is likely to be small.

[18] This assumption is more likely to hold for the serious violence measures that are our focus, rather than for lesser offenses where there is more room for police discretion; see Appendix A.2.5.

understate treatment effects.

3.2 Primary outcomes

To proxy for the latent variable of interest, we pre-specified a single primary outcome, *Serious violence involvement*, which is an index that standardizes and averages arrests and reported victimizations for serious violent crimes over the 20 months post-randomization.[19] The three components of this index are: (1) *Shooting and homicide victimizations*, (2) *Shooting and homicide arrests*, and (3) *Other serious violent-crime arrests*.[20]

Though READI's emphasis is on reducing shootings and homicides, we included arrests for serious but not gun-related violent offenses in the primary outcome. We did this because we initially underestimated READI's ability to identify people with future shooting involvement, and we were worried a sole focus on shootings would leave too few incidents to detect program effects. Similarly, in our pre-analysis plan we pooled shooting and homicide arrests and other serious violent-crime arrests into a single component, as we did not anticipate having sufficient power to detect effects on shooting and homicide-specific outcomes. However, upon seeing the high level of risk among men in the study sample, and given READI's focus on preventing shootings and homicides, we opted to separate these incidents from arrests for other serious violent crimes when constructing the index. This decision has a negligible impact on our estimates of READI's effect on the index itself (Appendix Table A.IV). When estimating effects separately for the three (rather than two) components of the index, the multiple hypothesis adjustment (described below) penalizes our inference for the additional hypothesis test.

An index is useful for increasing power to the extent that all underlying components move in the same direction—something we initially expected. But because READI can affect index

[19] The 20-month follow-up was chosen to measure effects during programming: 2 months to locate and recruit men, plus 18 months of possible programming. We plan a 40-month analysis as well.

[20] Arrests are limited to incidents that occurred during the 20-month outcome period. Shooting arrests include any aggravated assault or aggravated battery that involves shooting a gun. Other serious violent-crime arrests include other, non-shooting or homicide violent offenses historically included in "Part I" of the Uniform Crime Reporting program: aggravated assault and aggravated battery (excluding homicide, manslaughter, and non-fatal shootings), robbery, and criminal sexual assault. See Appendix A.2.2.

components differently, and to better understand what behavior may be changing, we also pre-specified that we would: (1) estimate effects separately for each component, correcting for the increased probability of Type I error that comes from testing multiple hypotheses; and (2) calculate an index where all arrests and victimizations are weighted by their social cost. Viviano et al. (2021) argue that a social cost-weighted index may be a more helpful way to summarize effects and provides a sufficient statistic for policy-making.

We implement two multiple hypothesis testing adjustments. First, we control for the family-wise error rate (FWER) among the three index components by using a free step-down resampling method (Anderson, 2008; Westfall and Young, 1993).[21] Second, we increase power by allowing for some proportion of null hypothesis rejections to be false, controlling for the false discovery rate (FDR) (Benjamini and Hochberg, 1995). We report the q-value, an analog to an adjusted p-value, which reports the smallest proportion of false null rejections we would have to accept to reject the hypothesis under the FDR control procedure.

3.3 Qualitative data collection

Between August 2017 and March 2020, two of the investigators and a qualitative research team conducted across all sites: (1) 220 hours of formal field observation; (2) 16 focus groups with 90 percent of READI staff in Spring 2019; and (3) 23 semi-structured, recorded interviews with participants. In addition, we piloted a survey instrument in the Austin/West Garfield Park site with 66 participants in Winter 2020. These data were gathered to describe participants' experiences and perceptions before and after take-up, inform program design, and aid in interpreting impact estimates. Because these data were collected prior to impact analysis, they were used to explore ex ante hypotheses of interest rather than to understand ex post patterns or outcomes—a form of exploratory and inductive qualitative research known as "emergent themes" (Williams, 2008). Appendix A.8 describes these data, methods of analysis, and conclusions in more detail.

[21] We report the adjusted p-value under strong FWER control. When we apply these adjustments to our heterogeneity tests, we treat the primary outcome index across the three referral pathways as a family, and then treat the index components within each referral pathway as its own family.

3.4 Estimating treatment effects

We estimate intent-to-treat (ITT) effects via the ordinary least squares regression:

$$Y_i = \beta T_i + \lambda \mathbf{X}_i + \gamma_s + \varepsilon_i \tag{1}$$

where Y_i is the 20-month outcome for individual i; T_i indicates assignment to an offer of READI; $\mathbf{X}_i$ is a vector of pre-randomization arrest, victimization, incarceration, and demographic characteristics; and γ_s is a vector of randomization strata fixed effects.[22] We estimate heteroskedasticity-robust standard errors.

The ITT estimate represents the effect of having an offer to participate in READI. Given incomplete take up, the ITT will understate the effect of READI participation. We therefore also estimate the treatment effect on the treated (TOT) by using random assignment as an instrument for participation. We define participation as a binary indicator equal to 1 for those who attended the initial orientation and signed job paperwork. To provide a sense of the proportional change for compliers, we report control complier means (CCMs).[23]

3.4.1 Threats to internal validity

Differential mortality and incarceration One consequence of evaluating programs for high-risk samples is that death or imprisonment will censor our dependent variables, possibly in ways that are correlated with treatment assignment. For example, if treatment delayed the risk of a homicide, then treatment group men would have had more opportunities than control group men to be arrested or victimized. In some respects, this change in time availability is part of the treatment effect. Overall counts of incidents still reflect how much violence there

[22] For more on pre-randomization characteristics, deviations from the pre-analysis plan, and robustness to the inclusion of different covariates and alternatives to strata fixed effects, see Appendix A.5.1.

[23] Because selection into take-up is likely non-random, CCMs, rather than control means, are the appropriate benchmarks for interpreting the magnitude of IV estimates (Abadie, 2003; Kling et al., 2007). The CCM is the estimated outcome mean for control group men who would have taken up READI had it been offered to them (control compliers). If no control group men take up (no always-takers), then the CCM is calculated by subtracting the TOT estimate from the outcome mean among treatment group compliers. Since 3 control group men took up, we use the CCM calculation in Heller et al. (2017).

was among each group, which is why we made these counts our primary outcomes.

Still, if we are interested in whether READI changed violent behavior and not just the number of events, then such a shift in incapacitation could mask changes in outcomes conditional on being able to engage in one's normal activities. Recognizing this, we pre-specified that we would test for differential rates of incarceration and death. Although we do not find significant differences in the overall amount of time treatment and control group men are incapacitated, we still verify in Appendix A.5.2 that our main results are robust to adjustment for differential censoring.

Spillovers The ITT and TOT are always interpretable as the average difference between the treatment and control groups (or compliers) after READI was implemented. But interpreting them as estimates of the direct effect of being offered or receiving READI relies on the Stable Unit Treatment Value Assumption (SUTVA)—that one person's treatment status does not affect another's potential outcomes. Given how social crime and violence can be, and that gun violence in particular is often a response to others' behavior, it is plausible that SUTVA could fail due to such spillovers (see Appendix A.5.4).

In a separate project, Craig et al. (in progress) estimate social spillovers. By combining multiple measures of social networks at baseline with the variation in treatment exposure generated by randomization, they identify how an intervention spreads through social networks while overcoming the typical challenges of endogenous ties and common shocks in the peer effects literature. While that project is broader than READI, Craig et al. (2022) reports early READI-specific results to aid with the interpretation of the estimates reported in this paper. They show that within the study population, there is no definitive sign of adverse spillovers (if anything, the results in this study may mask declines in other serious violent-crime arrests). But given the sample size, the analysis is somewhat under-powered.[24]

[24] They find clearer evidence of a READI-driven decline in drug-crime arrests that is masked in the ITT estimate reported here, because control group men who have co-arrest ties to treatment group men also show reduced drug-crime arrests. Combined with declines in these arrests among the peers of treatment group men who were outside the study sample entirely, it is likely the ITT estimate reported in this paper misses a net decline in drug-crime arrests from the intervention.

4 Descriptive statistics, realized risk, and take-up

4.1 Baseline characteristics and balance

Table I reports baseline summary statistics and tests of balance. The average age of study men at referral was 25, and nearly all (97 percent) are Black. Summary statistics on past arrest and victimization records confirm that the three referral pathways successfully identified men with very high levels of prior violence involvement. Baseline counts cover since 2010 for shooting victimizations and since 1999 for other arrest and victimization outcomes. For every 100 men in the sample, there were an average of 7.6 prior arrests for shootings or homicides, 46 prior shooting victimizations, and 90 arrests for other serious violent crimes (aggravated assault and battery, sexual assault, and robbery). The risk score shows that, based on our algorithm, we expected 11.4 percent of the sample to be either arrested for or the victim of a violent gun crime in the 18 months after referral.

Prior involvement in other kinds of crime and violence was also high. The average study member experienced over 17 arrests prior to randomization, 3.4 reported victimizations, and had spent roughly 175 days in jail or prison in the prior 30 months, with 4 percent being incarcerated at the time of randomization. Though we lack any direct way to measure it with the data available to us, prior ethnographic work (e.g., Aspholm, 2020) and the observations of READI program staff suggest that many of these men were involved in the neighborhood cliques that make up Chicago's fractured gang landscape.

Of the 17 baseline variables in Table I, two have treatment–control differences with $p < 0.1$—no more than would be expected by chance. A joint test of significance has a p-value of 0.3. We control for these variables in our analysis to account for any chance differences.

We also describe the limited information we have on our sample's educational attainment and employment before READI. Importantly, this information is only available for treatment group men who took Heartland Alliance's intake survey during orientation. Because of the selection into this sample, these data are not part of our balance tests or experimental analysis. Still, they are useful for painting a fuller picture of READI participants.

On average, 60 percent of the 535 survey respondents reported having less than a high school diploma. Thirty-one percent reported having graduated or earned a GED, and 8 percent reported attending some college without a degree.

Heartland's survey question about employment changed over time. For the 207 people with data before February 2019, 52 percent reported working in the last year, and just under 12 percent reported having never worked. The 305 people with data after that point were asked only about working "prior to READI," without a specific look-back period. Twenty-eight percent reported working for wages; 15 percent reported being self-employed; and 55 percent reported being out of work or unable to work. The Heartland survey did not explain whether participants should consider work in the illegal market as "working." In a survey we conducted of 66 participants in Austin/West Garfield Park, 45 percent reported receiving a formal paycheck in the last typical week before READI, consistent with the results of the Heartland survey and suggestive of respondents reporting mostly legal work. In addition, 38 percent of respondents to our survey reported legal but informal work, referred to as a "side hustle" that did not require risk. Finally, 59 percent reported a risky side hustle, a proxy for illegal work, including drug sales and theft.

4.2 Pathway differences and realized risk

The top panel of Table II reports baseline statistics by pathway (Appendix Table A.I shows baseline balance by pathway). Relative to outreach referrals, algorithm referrals had longer arrest and victimization histories on most measures, despite being about a year younger. Their number of prior serious violence incidents was 25 to 119 percent larger, depending on the measure, with a predicted level of future gun violence involvement of 14 versus 9 percent. Algorithm referrals also had about twice the number of prior reported victimizations. The only measure on which algorithm referrals looked less involved in crime is on days incarcerated in the past 30 months (129 versus 155 days), possibly because being incarcerated during the baseline period likely reduced the number of recently observed incidents, and hence the potential risk score. In general, re-entry referrals were in between algorithm and outreach referrals, though they were about a year older at baseline than the latter. These observable differences across referrals suggest that the pathways identified different kinds of people.[25]

Because a core question about the referral mechanisms was whether they could anticipate future risk, the bottom panel of Table II shifts from baseline characteristics to realized risk in the control group over the 20-month outcome period. It shows that READI's referral pathways successfully identified men at high risk of future arrest and victimization. Almost two-thirds of the control group were arrested during the 20 months after randomization, 1.7 times on average. About a third reported at least one victimization. The best-measured and most severe indication of gun violence involvement—being shot or killed—is shockingly high: there were 11 shooting and homicide victimizations per 100 control group members during the 20-month follow-up period. This is 52 times higher than the rate among average Chicagoans (0.21 per 100), and 2.7 times higher than the rate among other men 18–34 living in the same neighborhoods READI serves (4.1 per 100).[26] In short, READI's referral pathways identified a group at immensely high risk of being shot or killed.

The algorithm was also successful at predicting a broader measure of gun violence. It was trained to predict involvement in a violent gun crime as an arrestee or a victim in the next 18 months (the table's last row). About 16 percent were actually arrested for or reported being victims of a violent gun crime in that period, higher than the 11.4 percent predicted at baseline. This could partly be due to rising violent crime rates city-wide during the outcome period relative to the data on which the algorithm was trained, and partly from the role of unobservables discussed below.

On general measures of criminal legal involvement such as overall arrests and victimizations, the realized risk levels of referrals from the three pathways mirror the significant

[25] Consistent with this is the fact that referrals from one pathway rarely included people who had previously been referred via another pathway; for example, only 35 initial outreach referrals had already been randomized via the algorithm pathway at the time of referral.

[26] Rates calculated based on shootings and homicides in Chicago from READI's launch (August 2017) through the end of the last study member's 20-month outcome period (November 2021) and population data from the American Community Survey.

differences in their baseline characteristics (top panel of Table II). Algorithm referrals were more likely to be arrested and victimized on both the extensive and intensive margins than outreach referrals, who in turn had higher rates and counts than re-entry referrals. Yet on measures of gun violence specifically, men identified across the three pathways did not significantly differ on realized risk, despite their significant differences in predicted risk.

Realized rates of gun violence involvement were similar across the pathways in the control group despite the differing risk levels predicted by observables. This suggests that staff using on-the-ground knowledge were leveraging unobservables that predict gun violence ($\hat{Y}(0)$). A key question, given the similarity of gun violence involvement across pathways, is whether the human decision-makers were also selecting on expected gains from program participation (β), and not just levels of the outcome. We address this in Section 5.3.

4.3 Take-up and participation

Given the extraordinary risk of violence in the study population, their interest in participating in a program with rules and restrictions was not self-evident. Table III reports rates of participation among men in the treatment group, overall and by pathway. Among all men randomized to READI offers, 55 percent started the program (defined as attending orientation). This take-up rate is comparable to interventions working with much less disconnected populations, such as teenage boys still attending Chicago public high schools in similar neighborhoods (Heller et al., 2017).

As expected, take-up was highest (78 percent) among outreach referrals. These men often already knew some program staff and had been screened on an interest in, and "readiness" for, programming (see Appendix A.8.5). Recruiting re-entry referrals was also relatively successful, with take-up of 60 percent. Fewer of the algorithm referrals actually participated (37 percent). Based on staff interviews and field observations, a primary reason is that outreach workers were unable to locate many algorithm referrals. Unfortunately, the outreach organizations did not keep formal and complete records, so we do not know exactly how many were unreachable. Once found, algorithm referrals were sometimes less trustful of outreach workers, or had other reasons to decline participation (such as already being employed, or now living in a distant neighborhood; see Appendix A.3.1).

The rest of the table breaks out hours by activity and total earnings conditional on take-up, reflecting participation within the 20-month outcome period. The first few columns report the distribution of work hours from the payroll data from Heartland Alliance (the employer of record for READI participants), which shows a positive skew in participation (mean hours worked, 560 hours, are greater than median hours worked, 437 hours). Due to incompleteness in the CBT and training attendance data requiring us to use extrapolation and summing across periods, we can only report average non-work participation (159 hours). Our estimates suggest that participants earned an average of about $9,650.

Because people took up the program after different lengths of time, and because in-person work paused in March 2020, it is difficult to read retention rates directly from Table III. Appendix Table A.II reports the average maximum possible hours worked for each group during the 20-month outcome period pre-COVID, based on how long it took them to begin participating. Some of the differences in hours across pathways in Table III reflect differences in time to take-up; in fact, prior to the pandemic, participants from all three pathways worked between 28 and 31 percent of their possible post-take-up hours during the outcome period. Appendix Table A.III details participation during and after the pandemic.

Figure II reports two measures of job retention, weekly from the time participants first attended orientation. The first, shown by the solid line, is the proportion of participants who worked at least one day after the time noted on the x-axis, as measured through payroll data. The top panel shows that, in the first few weeks after orientation, roughly 15 percent of participants stopped returning to work. Afterwards, the decline in participation becomes roughly linear and relatively slow. About 75 percent of participants continued showing up to work after the 20-week mark, and a little over half continued to work after one year.[27]

[27] Figure II excludes the period starting in March 2020 when READI's in-person employment programming was suspended for several months. A version of the plot including this period shows similar patterns (Appendix Figure A.III), but with slightly lower participation towards the end of the program period.

The second measure, shown by the dotted line, is the proportion of weeks worked by participants who are still working. Unlike the first measure, which captures the extensive margin (*whether* people still work), this second measure captures the intensive margin (*how much* people work). After the initial fall-off, active participants consistently worked about 75 percent of the weeks that they could have.

The bottom panel of Figure II shows retention by pathway. (This excludes re-entry since the Figure is limited to pre-COVID data and too few re-entry participants started this early.) The overall patterns are quite similar for algorithm and outreach participants, with the key difference being a faster fall-off in early participation among algorithm participants. This resulted in the larger dosage for outreach participants shown in Table III.

5 Results

5.1 Average treatment effects

Table IV reports estimates of average treatment effects on our pre-specified primary outcome index and its components. The point estimate for being randomly assigned a READI offer (the ITT) is a 0.027 standard deviation reduction in the serious violence index (a 0.049 standard deviation reduction for the effect of participating, the TOT), but the result is not statistically significant ($p = 0.26$).

The second panel shows results on the three components of the index. The two measures of shooting and homicide involvement—arrests and victimizations—have negative and substantively large point estimates. There are 1.3 fewer shooting and homicide victimizations for every 100 participants, a 12 percent decline relative to the control complier mean. But the confidence interval is too wide to rule out a similarly-sized increase instead. The decline in arrests is a proportionally huge 65 percent (2.2 fewer per 100 participants). It is statistically significant on its own, but not after adjusting for the three hypothesis tests across components (unadjusted $p = 0.05$, adjusted $p = 0.13$).[28]

[28] We focus on reporting adjusted p-values that strongly control the FWER in the text, which is a conservative inference adjustment. The tables also report the less-conservative q-values from controlling the FDR.

The small reduction in the overall index reflects the fact that not all forms of violence move in the same direction. There are 0.6 additional arrests for other serious violent crimes for every 100 participants, an 11 percent increase, though the standard errors are large relative to the point estimate.[29] As we show in Section 6, however, the disproportionately high costs of shootings and homicides means that weighting arrests and victimizations by their social cost leads to different conclusions, which is relevant for policymakers to consider along with the lack of clear change in our primary outcome.

These results remain similar when using: a count model; different (or no) baseline covariates; randomization inference; adjustments for the possibility that incarceration or death are generating censoring (with some indication of a decline in the rate of shooting and homicide victimization that is masked by the additional days that treatment group men are alive); and the pre-COVID period only (see Appendix A.5). We also find that there are few statistically significant changes in other types of arrests, victimizations, or incarceration outcomes. Lastly, Appendix A.5.5 shows how treatment effects accrue over time.

A natural question is whether the large decline in shooting and homicide arrests is a direct result of incapacitation from the program itself—whether keeping people busy during the workday mechanically reduced violence during that time. Appendix A.5.6 reports estimated effects on arrests and victimizations separately by day and time, as measured by the time of incident (not the time of arrest). While point estimates on the total number of arrests and victimizations are negative and substantively large during work days, the fall in incidents underlying shooting and homicide arrests happens during weekends, suggesting that incapacitation does not seem to be driving the change.[30] This is corroborated by interview data, in which participants report changing with whom and where they spend time outside of READI hours (see Appendix A.8.6).

[29] Because the index components do not all move in the same direction, using the first component of a principal component analysis affords a bit more precision than an unweighted index: ITT $= -0.0569$, SE $= 0.0377$, $p = 0.13$. Because this approach was not pre-specified, we do not emphasize these findings.

[30] Appendix A.5.7 shows that there may be some role for incapacitation in reducing arrests for drug crimes.

5.2 Heterogeneity analysis

Here we examine heterogeneity of impacts by pre-specified subgroups. Because they were built into the experimental design and prediction process, we focus on differences across referral pathways and predicted risk levels.[31] Our pre-analysis plan noted that these tests would be exploratory, as we did not anticipate being powered to detect moderate heterogeneity. Nonetheless, we still adjust our inferences for multiple testing.

Table V shows significant heterogeneity across referral pathways: we can reject the null that the estimated effects on the primary outcome index are the same across pathways ($p = 0.03$). There is a clear, statistically significant decline in serious violence within the outreach pathway. The decline in the index of 0.13 standard deviations for outreach participants remains statistically significant after adjusting inference for the three tests across pathways (adjusted $p = 0.03$). Breaking the index into components shows a similar but more precise pattern as in the overall results: large declines in both arrests (79 percent) and victimizations (43 percent) for shootings and homicides, which both remain statistically significant after adjusting for the three tests across outcomes (adjusted $p = 0.03$ and 0.08, respectively), with a small decline (but large standard error) on other violent-crime arrests.

No results in the other referral pathways approach statistical significance. In many cases this is because the tests are under-powered, not because point estimates are substantively small. Standard errors for the TOT in the algorithm pathway are particularly large, despite the larger sample there, because of the weaker first stage resulting from lower take-up.

Table VI shows that pathway differences are not driven by differences in baseline predicted risk. The table reports separate effects for those over and under median predicted risk, as well as the 231 people with missing scores.[32] There are some indications that READI's

[31] Appendix A.6 summarizes heterogeneity by two other pre-specified subgroups, site and age. Briefly, we find no evidence of heterogeneity by age, but there is some variation by site: significantly larger declines in two sites (Austin/West Garfield Park and Greater Englewood) but a positive point estimate in the third (North Lawndale).

[32] Missing a risk score (which can only happen for outreach and re-entry referrals) indicates that someone had too little recent police contact to meet the criteria for inclusion in the algorithm: at least one arrest or two victimizations in the last 50 months. Consistent with the idea that less recent police contact indicates

effects may be concentrated among those with higher predicted risk: the point estimate on the primary index is large and negative only for that group, and the only statistically significant estimate among the components is the decline in shooting and homicide arrests for those at above-median predicted risk. Overall, we cannot reject the null that the three predicted risk groups have the same effect ($p = 0.33$), and the confidence interval is wide enough to include large proportional declines from a lower baseline for the below-median predicted risk group.

5.3 Interpreting pathway heterogeneity

Motivated by the patterns in the previous section, we explore outcomes by pathway and predicted risk, focusing on how the outreach workers' selection process or role in participant engagement might explain the observed treatment heterogeneity. Figure III presents all the results we discuss in this section. It shows outcomes by pathway and baseline risk score quartile. The quartiles are defined across the full distribution of risk scores in the study sample, so average predicted risk within each quartile is very similar across pathways.[33]

The top panel of Figure III shows ITT estimates for the primary outcome index (see Appendix Figure A.VI for components). For outreach referrals with below-median or missing predicted risk, the point estimates are very close to zero. The large estimated declines in serious violence for outreach referrals in Table V are driven by the smaller group of above-median predicted risk outreach referrals. Yet there do not appear to be parallel declines among the above-median predicted risk algorithm referrals. Thus, there seems to be an interaction between the predictable part of the risk of gun violence involvement and something outreach workers are doing.

The two bottom panels of Figure III help rule out several possibilities about how outreach workers matter. The bottom left panel shows average rates of realized gun violence

lower risk, the control means for the missing group are considerably lower.

[33] For distributions of risk scores within referral pathway, see the top left panel of Appendix Figure A.VI. While we describe referrals in quartiles 1 and 2 as having "lower" predicted risk, in absolute terms it is still quite high: almost 7 percent of this group was expected to be arrested for or the victim of a violent gun crime within 18 months, compared to 11 percent for the full sample.

involvement (the outcome predicted by the algorithm) by pathway and risk quartile for the control group (i.e., $\bar{Y}(0)$). The bottom right panel shows take-up and hours worked for the treatment group. Marker sizes correspond to the share of referrals within each pathway in each quartile.

The marker sizes in both bottom panels show that outreach workers did not prioritize the highest predicted risk referrals: most outreach referrals (75 percent) had below-median or missing predicted risk, while most algorithm referrals (67 percent) had above-median predicted risk. The fact that most outreach referrals had lower predicted risk—not the subgroup most responsive to READI—suggests that outreach workers were not consistently successful at selecting on gains, even if they were trying to do so.

The bottom panels suggest two other factors with a stronger relationship to outreach referral decisions: the unobservables that contributed to a high risk of gun violence and the proclivity to take up the program. Outreach workers' success at identifying unobservable risk factors is shown in the bottom left panel. While we know from Table II that both pathways referred men whose realized risk was greater than their predicted risk, outreach referrals typically had higher realized risk than algorithm referrals at similar predicted risk levels. The gap between the pathways suggests that outreach workers selected men partly based on information unobservable to the algorithm that is predictive of high $Y(0)$. The bottom right panel shows that outreach referrals were about 40 percentage points more likely to start READI, regardless of predicted risk. This pattern is consistent with outreach workers referring men partly based on the unobservable (to the researcher) determinants of take-up, potentially including the ability to find the referrals in the first place.[34]

Importantly, however, neither factor appears to explain the pattern of treatment hetero-

[34] This possibility was also highlighted in the qualitative data, where staff frequently discussed selecting on "readiness for READI" (see Appendix A.8.5). Examples of not being ready offered by outreach workers included men who were still focused on settling a score, men who preferred illegal work (such as drug dealing), men who feared for their safety if they left their block or associated with opposition members, men who could not overcome their impatience with or skepticism of the therapy and jobs, and men unwilling to attend the program sober. While the concentration of outreach referrals in the less responsive, lower predicted risk quartiles suggests that being "ready for READI" does not necessarily mean ready for change, it could still mean ready to take up the program.

geneity. The gap in realized risk between algorithm and outreach referrals is similar across quartiles 2 and 3, and if anything, slightly larger in quartile 2. But the estimated treatment effect is much larger for outreach referrals in quartile 3. Although this is not a strong test of the role of unobservable determinants of $Y(0)$, the pattern seems inconsistent with the idea that outreach workers' use of such unobservables in their referral decision is the key driver of the larger effects among that pathway.[35] Moreover, take-up rates are fairly flat across the risk quartiles.[36] Because treatment effects are concentrated in the top risk quartiles, this argues against selection on take-up being the sole driver of pathway treatment heterogeneity.

Another possibility is that variation in treatment dosage drives treatment heterogeneity, and dosage is positively correlated with predicted risk. To assess this possibility, the numbers above each marker in the bottom right panel show participation decisions on the intensive margin (hours worked during the 20-month outcome period). Interestingly, while outreach referrals were more likely to come from the lower predicted risk quartiles that had higher hours worked, outreach participants did not consistently work more hours after starting than did algorithm participants in the same quartile.[37] Nor did outreach participants in the top two risk quartiles consistently work more hours. So it does not appear that the intensive margin of participation explains the patterns of treatment heterogeneity on its own, either.

Appendix A.6.5 provides some evidence that the unobservables correlated with pathway, rather than dosage itself, matter for treatment heterogeneity.[38] The fact that algorithm

[35] Another possibility is that outreach workers are selecting on observables, but in a different way than is captured in the algorithmic risk score. Appendix Sections A.6.2 and A.6.3 provide evidence that this is not the case. Baseline covariates only weakly predict whom the outreach workers will select (Appendix Table A.XIV), and adjusting algorithm referrals' observables to look like those of outreach referrals does little to close the gap in estimated treatment effects between these groups (Appendix Table A.XV).

[36] Note that relatively flat take-up rates across quartiles in each pathway also mean that the patterns for the TOT estimates are similar to the ITTs we show in the top panel.

[37] Appendix A.6.4 expands on this pattern by showing that pathway explains more of the variation in take-up than the other observables, but little of the variation in hours worked beyond what observables explain.

[38] In Appendix A.6.5, we adapt the Abadie et al. (2018) endogenous stratification exercise to predict dosage among controls, then estimate heterogeneous treatment effects for low- versus high-predicted dosage groups. We find suggestive evidence that treatment effects are larger for the high-predicted dosage group. But when we do the stratification exercise without pathway as a dosage predictor, we do not find the same pattern. This is further support for the idea that the unobservables correlated with pathway rather than dosage itself matter for treatment heterogeneity. It may be, then, that the kinds of people who select into

participants often worked as much as outreach participants in the same risk quartile is also suggestive evidence that outreach workers did not just generate a more supportive environment for their own referrals, or work harder to keep them in the program. Of course, hours worked is not the only measure of dosage, so we cannot rule out differential dosage across pathways. But we also found little evidence in our interviews and focus groups that outreach workers treated participants differently based on their pathway after take-up.

In the end, while we have some evidence for what outreach workers used in their selection process (unobserved predictors of $Y(0)$ and take-up), we can rule out that these factors alone explain the larger responsiveness among outreach referrals. That leaves us without a clear explanation for the patterns of treatment effects by pathway and predicted risk of gun violence involvement. It is important for future research to further explore why this interaction occurs, as different possibilities have different implications for policy.

For example, it is natural to look at the pattern in the top panel of Figure III and conclude that while outreach workers are essential, screening out referrals with lower predicted risk might be beneficial. This would be true if those with lower $\hat{Y}(0)$ are genuinely less responsive, including the possibility of a floor effect with too little room for decreases in violence (Heller et al., 2022). If so, continuing to have outreach workers make referrals, or finding other ways to identify people with the relevant unobservables, would remain crucial for reducing violence. And continuing to make referrals solely from an algorithm trained to predict $\hat{Y}(0)$ would be ineffective. The combination of outreach screening and the algorithm would still be important, though, both to limit outreach referrals to those most responsive and to bring those whom outreach workers might not have otherwise considered into the screening process.

But two possibilities might change those policy implications. First, it is possible that READI staff had an easier time finding alternative programming for control group men with lower predicted risk since outreach workers already knew them—even though, as discussed earlier, programming comparable to READI was seldom available. If so, then the apparent

high doses of READI also respond more, but for reasons other than a direct dose-response relationship.

lack of responsivity could be due to an improved counterfactual rather than a lower potential gain from READI. If access to counterfactual programming matters, then screening out people with lower predicted risk would only be productive in settings with enough access to outside programming. Second, outreach workers could have engaged more successfully with their own referrals or generated a more productive environment for them, but in a way that did not consistently result in outreach participants working more hours than algorithm participants in the same risk quartile. The role of ongoing interaction and relationships with caseworkers seems to play a role in program success in other settings, such as sectoral employment training programs (see, e.g., Katz et al., 2022). If such personal engagement is key, then it might not be necessary or desirable to exclude higher predicted risk algorithm referrals from programming: efforts could instead focus on improving relationships and increasing engagement with such referrals to raise their responsiveness to treatment.

We emphasize that the findings in this section are exploratory: our sample is too small to determine whether all of these subgroup results differ from each other significantly. On their face, they suggest that the combination of human and algorithmic screening used by READI seemed to outperform either referral mechanism individually, which may provide some guidance for future CVI design and for research. But future work confirming this heterogeneity, and unpacking the reasons for it, would be useful.

6 Social benefits and costs

In many settings, the statistical uncertainty about declines in primary outcome components, combined with the lack of improvements in other secondary outcomes, might lead us to question the value of the intervention. But there is a crucial difference about this setting: here, the outcomes of interest are literally measures of life and death. Shootings and homicides generate such enormous costs to individuals, families, and communities that even limited improvement on a small number of outcomes could generate large benefits to society. Weighting outcome measures by their social value can help inform how to think about the value of the program, and can provide a useful alternative to multiple testing adjustments when policymakers value changes in the different outcomes differently (Viviano et al., 2021).

Here we try to quantify READI's benefits and then compare them to program costs. Our main focus is on the social costs surrounding crime and violence. Our calculations are necessarily a rough approximation, both because of the uncertainty involved with assigning costs to crime and the loss of life (Dominguez and Raphael, 2015), and because we ignore a range of other difficult-to-measure costs and benefits (e.g., the social value of investing and generating jobs in historically under-served neighborhoods, the value of the work that READI participants do, gains in unmeasured outcomes like mental or physical health, and the opportunity cost of program spending).[39] Our intention is not to provide a definitive cost-benefit analysis, but rather to provide a basic sense of how the substantive importance of the outcomes READI affects might shape how we think about the program's impacts.

Because of uncertainty about the social costs of crime (including the statistical value of life), and because there are numerous choices about how to translate imperfect measures like arrests into assessments of the true amount of crime, Table VII presents sets of less and more inclusive estimates for how READI affects social costs. Since there is only one outcome and the analysis was pre-specified, we do not adjust for multiple testing.

We assign to each arrest and victimization of a READI study member an estimated cost of crime depending on its type. Appendix A.7 contains the details of these calculations. The less inclusive estimates use the lower-end estimates of the cost of crime from Cohen and Piquero (2009), and only count the harm from crime when directly observed in the data (i.e., from each observed arrest or victimization of a READI study participant). The more inclusive estimates use the higher willingness-to-pay estimates of the cost of crime (Cohen and Piquero, 2009); extrapolate how much crime is likely to have occurred for every arrest, given clearance rates; and scale each victimization up by average reporting rates for each

[39] Because READI was largely funded by private philanthropy, there was no deadweight loss from raising tax dollars. Future versions of the program might use public funding, and so might benefit from weighing the marginal value of public funds (Hendren and Sprung-Keyser, 2020).

crime type.[40] When we compare the benefits of reduced crime to the administrative costs of the program, we treat program wages as a transfer from society to participants. The net READI costs are thus the administrative costs less payments to READI participants.

Table VII reports program impacts on these costs. The first thing to note is how much cost society incurred from the level of crime and violence in READI's absence: between \$360,000 and \$1,870,000 on average, depending on how inclusive the costs are, for each control complier. The second thing to note is that despite the uncertainty about how each element of our primary index responds to READI, the severity of shootings and homicides is so great that their higher weights in the cost of crime calculations generate less uncertainty about program benefits. We estimate that READI saved society a total of between \$182,000 and \$916,000 for each participant ($p = 0.03$ and $p = 0.02$, respectively), about a 50 percent decline. Compared to the cost of offering READI over 20 months (about \$52,000 for each participant), these benefits are at least 4—and perhaps as much as 18—times as large as the program's costs.

7 Discussion

The READI study was designed to determine (1) whether it is possible to find people at high enough risk of gun violence for a targeted intervention to reduce shootings sufficiently to outweigh program costs; (2) whether they could be engaged; and (3) whether the combination of jobs, CBT, and outreach could reduce their involvement in serious violence. The answer to the first two questions is a definitive yes. Each referral pathway found men at extraordinary risk of being shot, killed, or arrested for a serious violent crime. Moreover, most of these men were willing to engage, despite the risks and barriers they faced. The 55 percent take-up rate for READI was higher than the take-up rates among high school boys for the in-school program studied by Heller et al. (2017) and among former prisoners for shorter transitional jobs programs (e.g., Redcross et al., 2016). Our interviews with participants suggest that

[40] In particular, for each crime type's clearance rate C, we assume there are $1/C$ crimes committed for each observed arrest, and C offenders arrested for every observed victimization. We use Chicago-specific clearance rates for homicides and clearance rates from Belfield et al. (2006) for other offenses.

many were initially motivated by the paycheck but stayed due to how the combination of CBT-related skills and caring program staff improved their well-being.

READI's impacts on serious violence, however, are mixed. There is no statistically significant decline in our pre-specified primary outcome. This means we cannot conclude with certainty that the version of READI evaluated here decreased serious violence. But there is suggestive evidence that READI reduced arrests for shootings and homicides, with the estimated effect being just beyond traditional statistical significance cutoffs. Moreover, for one subgroup—men referred by outreach workers—the declines in arrests and victimizations for shootings and homicides clearly pass standard statistical significance thresholds.

If program operators are limited to one referral pathway, using only outreach workers to identify referrals may therefore be the best way to maximize treatment effects. This is important for external validity, since jurisdictions will likely find it easier to implement an outreach referral pathway; many cities already have local organizations working with people at high risk of gun violence. In contrast, it may be more difficult to gain access to detailed police or other administrative data useful for predicting gun violence risk, find the expertise to train an algorithmic model, and convince local providers to trust the results.

Still, our exploratory analysis shows that only the outreach referrals with the highest algorithmic risk predictions respond to READI. So it is possible that a combination of human intelligence and machine-driven risk prediction may more effectively anticipate treatment responsiveness than either method alone. The fact that men with equally high risk of gun violence seem to respond quite differently depending on how they were selected for READI highlights the importance of future research on what explains the interaction between predicted risk and referral pathway: differences in selection by outreach workers, in program experiences by pathway, in responsivity by risk level, or in varying counterfactuals.

In considering external validity, several findings echo patterns in other violence prevention studies. Despite differences in setting, population, and treatment between READI, a CBT program serving former members of armed groups in Liberia, and summer jobs programs across several U.S. cities, both Blattman et al. (2022) and Heller (2022) find larger effects for those facing the largest risk of negative outcomes. And a completely different violence-prevention program, which involves information sharing between hospitals, local government, and police, finds a decline in the most lethal outcomes combined with an increase in less lethal violence, similar to our pattern of results (Florence et al., 2011). It will always be important to attend to context and diagnosis of the conditions driving violence. The blend of therapy and economic engagement seems to work in places like Chicago and Liberia, where a significant fraction of the violence arises from fragmented groups who are responding to perceived slights, injustices, and reputational concerns; it may work less well in contexts where killings are more rooted in organized crime or instrumental violence (Blattman, 2022). Still, commonalities such as variation in responsiveness by risk level (at least among some subgroups) and substitution from more to less lethal violence raise the possibility that similar kinds of behavior change are possible across settings.

Our estimates do not capture READI's complete impact on participants owing to several measurement challenges. For example, while crimes like aggravated assault and robbery are far more common than shootings and homicides, reported victimizations for the latter are twice as common as arrests for the former in our data, driven entirely by differences in measurement. Arrests outside of Chicago also remain unobserved. And administrative data will always yield an incomplete view of a program's impact. As one example, our qualitative data collection made clear that the investments in the community (both the urban renewal work participants performed and the READI-induced hiring at local organizations) were valuable to those who took part in READI. At least some participants reported unmeasured impacts on relationship quality, self confidence, and community integration.

From a scientific perspective, the statistically insignificant decline in our primary outcome merits caution. Science is rightfully conservative about overturning null hypotheses based on a single imprecise finding. From a policy perspective, however, the binary conclusion from one hypothesis test is not dispositive; $p > 0.1$ is not the same as a true zero. As a number of economists have emphasized (Imbens, 2021; Manski, 2019; Ziliak and McCloskey, 2008), policymakers should attend to the range of plausible magnitudes consistent with a point estimate, along with the availability of other credible evidence and the benefits and costs of the status quo versus new investments. In this vein, it is worth emphasizing that about 74 percent of the confidence interval for READI's estimated impact on the serious violence index is below zero; 87 percent of the confidence interval for shooting and homicide arrests is negative. Weighting arrests and victimizations by their social costs shows that READI's estimated benefits to society were at least 4 times its cost.

For policymakers trying to reduce the enormous harms of gun violence, and given the absence of other convincing evidence about how to reduce shootings without the collateral costs that often accompany aggressive law enforcement, the findings we report here should still be useful inputs into decision-making, despite the need for more evidence to generate firm scientific conclusions. READI is a demonstration that investments of this scale are possible, even among a population deeply disconnected from every institution other than the criminal legal system, with a potentially large social payoff given the concentration of socially costly outcomes among a relatively small group of people. But future research could usefully experiment with refinements to the program model, targeting strategies, and scaling challenges, and other ways to reduce the extraordinarily high risk of gun violence among a small and under-served group.

8 References

Abadie, Alberto (2003) "Semiparametric instrumental variable estimation of treatment response models," *Journal of econometrics*, 113 (2), 231–263.

Abadie, Alberto, Matthew M Chingos, and Martin R West (2018) "Endogenous stratification in randomized experiments," *Review of Economics and Statistics*, 100 (4), 567–580.

Abt Associates (2021) "Implementation Evaluation of Roca Inc.," `https://www.abtassociates.com/files/Projects/PDFs/2021/final-report_abt-associates_roca-implementation-evaluation.pdf`.

Abt, Thomas (2019) *Bleeding out: The devastating consequences of urban violence–and a bold new plan for peace in the streets*: Hachette UK.

Anderson, Elijah (1999) *Code of the street: Decency, violence, and the moral life of the inner city*: WW Norton & Company.

Anderson, Michael L. (2008) "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103(484), 1481–1495.

Andrews, Isaiah and Emily Oster (2021) "a simple approximation for evaluating external validity bias," National Bureau of Economic Research.

Ang, Desmond (2021) "The Effects of Police Violence on Inner-City Students," *Quarterly Journal of Economics*, 136 (1), 115–168.

Arbour, William (2022) "Can Recidivism Be Prevented From Behind Bars? Evidence From a Behavioral Program."

Aspholm, Roberto (2020) *Views from the streets: The transformation of gangs and violence on Chicago's south side*: Columbia University Press.

Beck, Aaron T (1979) *Cognitive therapy of depression*: Guilford press.

Beck, Aaron T and David JA Dozois (2011) "Cognitive therapy: current status and future directions," *Annual review of medicine*, 62, 397–409.

Belfield, Clive R, Milagros Nores, Steve Barnett, and Lawrence Schweinhart (2006) "The high/scope perry preschool program cost–benefit analysis using data from the age-40 followup," *Journal of Human resources*, 41 (1), 162–190.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014a) "High-Dimensional Methods and Inference on Structural and Treatment Effects," *Journal of Economic Perspectives*, 28 (2), 29–50.

——— (2014b) "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 81 (2), 608–650.

Benjamini, Yoav and Yosef Hochberg (1995) "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, 57 (1), 289–300.

Berk, Richard, Lawrence Sherman, Geoffrey Barnes, Ellen Kurtz, and Lindsay Ahlman (2009) "Forecasting murder within a population of probationers and parolees: a high stakes application of statistical learning," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172 (1), 191–211.

Blattman, Christopher (2022) *Why We Fight: The Roots of War and the Paths to Peace*: Viking.

Blattman, Christopher, Sebastian Chaskel, Julian C Jamison, and Margaret Sheridan (2022) "cognitive behavior therapy reduces crime and violence over 10 years: Experimental evidence," National Bureau of Economic Research.

Blattman, Christopher, Donald P Green, Daniel Ortega, and Santiago Tobón (2021) "Place-based interventions at scale: The direct and spillover effects of policing and city services on crime," *Journal of the European Economic Association*, 19 (4), 2022–2051.

Blattman, Christopher, Julian C Jamison, and Margaret Sheridan (2017) "Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in Liberia," *American Economic Review*, 107 (4), 1165–1206.

Bloom, Howard S, Larry L Orr, Stephen H Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M Bos (1997) "The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act study," *Journal of human resources*, 549–576.

Braga, Anthony A (2003) "Serious youth gun offenders and the epidemic of youth violence in Boston," *Journal of Quantitative Criminology*, 19 (1), 33–54.

Braga, Anthony A and Philip J Cook (2023) *Policing gun violence: Strategic reforms for controlling our most pressing crime problem*: Oxford University Press.

Braga, Anthony A, David M Kennedy, Elin J Waring, and Anne Morrison Piehl (2001) "Problem-oriented policing, deterrence, and youth violence: An evaluation of Boston's Operation Ceasefire," *Journal of research in crime and delinquency*, 38 (3), 195–225.

Braga, Anthony A, Andrew V Papachristos, and David M Hureau (2014) "The effects of hot spots policing on crime: An updated systematic review and meta-analysis," *Justice quarterly*, 31 (4), 633–663.

Braga, Anthony A, David Weisburd, and Brandon Turchan (2018) "Focused deterrence strategies and crime control: An updated systematic review and meta-analysis of the empirical evidence," *Criminology & Public Policy*, 17 (1), 205–250.

Buggs, Shani A, Daniel W Webster, and Cassandra K Crifasi (2022) "Using synthetic control methodology to estimate effects of a Cure Violence intervention in Baltimore, Maryland," *Injury prevention*, 28 (1), 61–67.

Butts, Jeffrey A, Caterina Gouvis Roman, Lindsay Bostwick, and Jeremy R Porter (2015a) "Cure violence: a public health model to reduce gun violence," *Annual review of public health*, 36, 39–53.

Butts, Jeffrey A, Kevin T Wolff, Evan Misshula, and Sheyla A Delgado (2015b) "Effectiveness of the cure violence model in New York City."

Carr, Jillian and Jennifer L Doleac (2016) "The geography, incidence, and underreporting of gun violence: new evidence using ShotSpotter data," `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2770506`.

CDC (2020) "WISQARS (Web-based Injury Statistics Query and Reporting System)|Injury Center," `https://www.cdc.gov/injury/wisqars/index.html`.

Chalfin, Aaron (2015) "Economic costs of crime," *The encyclopedia of crime and punishment*, 1–12.

Chalfin, Aaron, Benjamin Hansen, Emily K Weisburst, and Morgan C Williams Jr (2022) "Police force size and civilian race," *American Economic Review: Insights*, 4 (2), 139–58.

Chandler, Dana, Steven D Levitt, and John A List (2011) "Predicting and preventing shootings among at-risk youth," *American Economic Review: Papers and Proceedings*, 101 (3), 288–92.

Cohen, Mark A and Alex R Piquero (2009) "New evidence on the monetary value of saving a high risk youth," *Journal of Quantitative Criminology*, 25 (1), 25–49.

Cook, Philip J, Ariadne E Rivera-Aguirre, Magdalena Cerdá, and Garen Wintemute (2017) "Constant lethality of gunshot injuries from firearm assault: United States, 2003–2012," *American journal of public health*, 107 (8), 1324–1328.

Corburn, J and A Fukutome (2021) "Advance peace Stockton, 2018-20 evaluation report."

Corburn, Jason, DeVone Boggan, Khaalid Muttaqi, and Sam Vaughn (2022) "Preventing urban firearm homicides during COVID-19: preliminary results from three cities with the Advance Peace Program," *Journal of urban health*, 1–9.

Craig, Ashley, Sara B Heller, and Nikhil Rao (2022) "A Preliminary Analysis of Spillovers in READI Chicago – Early Results from "Using Network Data to Measure Social Returns and Improve Targeting of Crime-Reduction Interventions"," `https://drive.google.com/file/d/1rbkj03yo_RAN2qdtjJFhhvS4WhcgApR2/view?usp=sharing`.

——— (in progress) "Using Network Data to Measure Social Returns and Improve Targeting of Crime-Reduction Interventions."

Cummings, Danielle and Dan Bloom (2020) "Can Subsidized Employment Programs Help Disadvantaged Job Seekers? A Synthesis of Findings from Evaluations of 13 Programs," *OPRE Report 2020-23*.

Dinarte, Lelys and Pablo Egaña del Sol (2019) "Preventing violence in the most violent contexts: Behavioral and neurophysiological evidence," *World Bank Policy Research working paper* (8862).

Dominguez, Patricio and Steven Raphael (2015) "The role of the cost-of-crime literature in bridging the gap between social science research and policy making: Potentials and limitations," *Criminology & Pub. Pol'y*, 14, 589.

Fagan, Jeffrey and Deanna L Wilkinson (1998) "Guns, youth violence, and social identity in inner cities," *Crime and justice*, 24, 105–188.

Farrell, Albert D, David Henry, Catherine Bradshaw, and Thomas Reischl (2016) "Designs for evaluating the community-level impact of comprehensive prevention programs: Examples from the CDC Centers of Excellence in Youth Violence Prevention," *The journal of primary prevention*, 37 (2), 165–188.

Farrington, David P, Jeremy W Coid, Louise Harnett, Darrick Jolliffe, Nadine Soteriou, Richard Turner, and Donald J West (2006) *Criminal careers up to age 50 and life success up to age 48: New findings from the Cambridge Study in Delinquent Development*, 94: Home Office Research, Development and Statistics Directorate London, UK.

Florence, Curtis, Jonathan Shepherd, Iain Brennan, and Thomas Simon (2011) "Effectiveness of anonymised information sharing and use in health service, police, and local government partnership for preventing violence related injury: experimental study and time series analysis," *Bmj*, 342.

Fox, Andrew M, Charles M Katz, David E Choate, and Eric C Hedberg (2015) "Evaluation of the Phoenix TRUCE project: a replication of Chicago CeaseFire," *Justice Quarterly*, 32 (1), 85–115.

Geller, Amanda, Jeffrey Fagan, Tom Tyler, and Bruce G. Link (2014) "Aggressive policing and the mental health of young urban men," *American Journal of Public Health*, 104 (12), 2321–2327.

Green, Ben, Thibaut Horel, and Andrew V Papachristos (2017) "Modeling contagion through social networks to explain and predict gunshot violence in Chicago, 2006 to 2014," *JAMA internal medicine*, 177 (3), 326–333.

Grogger, Jeffrey (2002) "The effects of civil gang injunctions on reported violent crime: Evidence from Los Angeles County," *The Journal of Law and Economics*, 45 (1), 69–90.

Harcourt, Bernard E (2005) *Illusion of order: The false promise of broken windows policing*: Harvard University Press.

Haveman, Robert, Rebecca Blank, Robert Moffitt, Timothy Smeeding, and Geoffrey Wallace (2015) "The war on poverty: Measurement, trends, and policy," *Journal of Policy Analysis and Management*, 34 (3), 593–638.

Heller, Sara B (2022) "When scale and replication work: Learning from summer youth employment experiments," *Journal of Public Economics*, 209, 104617.

Heller, Sara B, Benjamin Jakubowski, Zubin Jelveh, and Max Kapustin (2022) "Machine Learning Predicts Shooting Victimization Well Enough to Help Prevent It," *NBER Working Paper No. 30170*.

Heller, Sara B, Anuj K Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A Pollack (2017) "Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago," *The Quarterly Journal of Economics*, 132 (1), 1–54.

Hendren, Nathaniel and Ben Sprung-Keyser (2020) "A unified welfare analysis of government policies," *The Quarterly Journal of Economics*, 135 (3), 1209–1318.

Hureau, David M, Theodore Wilson, Wayne Rivera-Cuadrado, and Andrew V Papachristos (2022) "The experience of secondary traumatic stress among community violence interventionists in Chicago," *Preventive medicine*, 107186.

Imbens, Guido W (2021) "Statistical significance, p-values, and the reporting of uncertainty," *Journal of Economic Perspectives*, 35 (3), 157–74.

Jones, Nikki (2014) "'The Regular Routine': Proactive Policing and Adolescent Development Among Young, Poor Black Men," *New Directions for Child and Adolescent Development* (143), 33–54.

Kapustin, Max, Jens Ludwig, Marc Punkay, Kimberley Smith, Lauren Speigel, and David Welgus (2017) "Gun violence in Chicago, 2016," *University of Chicago Crime Lab*.

Katz, Lawrence F, Jonathan Roth, Richard Hendra, and Kelsey Schaberg (2022) "Why do sectoral employment programs work? Lessons from WorkAdvance," *Journal of Labor Economics*, 40 (S1), S249–S291.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz (2007) "Experimental analysis of neighborhood effects," *Econometrica*, 75 (1), 83–119.

Lipsey, Mark W, Nana A Landenberger, and Sandra J Wilson (2007) "Effects of cognitive-behavioral programs for criminal offenders," *Campbell systematic reviews*, 3 (1), 1–27.

Lochner, Lance (2004) "Education, work, and crime: A human capital approach," *International Economic Review*, 45 (3), 811–843.

Lochner, Lance and Enrico Moretti (2004) "The effect of education on crime: Evidence from prison inmates, arrests, and self-reports," *American economic review*, 94 (1), 155–189.

Loftin, Colin and David McDowall (2010) "The use of official records to measure crime and delinquency," *Journal of quantitative criminology*, 26 (4), 527–532.

Manski, Charles F (2019) "Treatment choice with trial data: Statistical decision theory should supplant hypothesis testing," *The American Statistician*, 73 (sup1), 296–304.

McNeill, Melissa and Zubin Jelveh (2022) "Name Match," `https://github.com/urban-labs/namematch`.

MDRC (1980) *Summary and findings of the national supported work demonstration*: Ballinger Publishing Company.

——— (2013) "Building Knowledge About Successful Prisoner Reentry Strategies," `https://www.mdrc.org/sites/default/files/Reentry_020113.pdf`.

Miller, Ted R (1996) *Victim costs and consequences: A new look*: US Department of Justice, Office of Justice Programs, National Institute of Justice.

Morgan, Rachel E and Alexandra Thompson (2021) "Criminal victimization, 2020," *Washington, DC: National Crime Victimization Survey, Bureau of Justice Statistics. Retrieved Jan*, 4, 2022.

Pattillo, Mary, Bruce Western, and David Weiman (2004) *Imprisoning America: The social effects of mass incarceration*: Russell Sage Foundation.

Picard-Fritsche, Sarah and Lenore Cerniglia (2013) *Testing a public health approach to gun violence: An evaluation of Crown Heights Save Our Streets, a replication of the Cure Violence Model*: Center for Court Innovation New York, NY.

Pyrooz, David, Jose Antonio Sanchez, and Elizabeth Weltman (2023) "Multidisciplinary Teams, Street Outreach, and Gang Intervention: Mixed Methods Findings from a Randomized Controlled Trial in Denver."

Raphael, Steven and Michael A Stoll (2013) *Why are so many Americans in prison?*: Russell Sage Foundation.

Redcross, Cindy, Bret Barden, Dan Bloom, Joseph Broadus, Jennifer Thompson, Sonya Williams, Sam Elkin, Randall Juras, Janaé Bonsu, Ada Tso et al. (2016) "The enhanced transitional jobs demonstration: Implementation and early impacts of the next generation of subsidized employment programs," *MDRC and the Employment and Training Administration, US Department of Labor, November*.

Ridgeway, Greg, Jeffrey Grogger, Ruth A Moyer, and John M Macdonald (2019) "Effect of gang injunctions on crime: A study of Los Angeles from 1988–2014," *Journal of quantitative criminology*, 35 (3), 517–541.

Roman, Caterina G, Hannah J Klein, and Kevin T Wolff (2018) "Quasi-experimental designs for community-level public health violence reduction interventions: a case study in the challenges of selecting the counterfactual," *Journal of Experimental Criminology*, 14 (2), 155–185.

Sherman, Lawrence W and Dennis P Rogan (1995) "Effects of gun seizures on gun violence:"Hot spots" patrol in Kansas City," *Justice Quarterly*, 12 (4), 673–693.

Skogan, Wesley G, Susan M Hartnett, Natalie Bump, Jill Dubois et al. (2008) "Evaluation of ceasefire-Chicago," *Chicago: Northwestern University*, 42 (5).

Sunstein, Cass R (2013) "The value of a statistical life: some clarifications and puzzles," *Journal of Benefit-Cost Analysis*, 4 (2), 237–261.

Viviano, Davide, Kaspar Wuthrich, and Paul Niehaus (2021) "(When) should you adjust inferences for multiple hypothesis testing?", `https://arxiv.org/abs/2104.13367`.

Webster, Daniel W, Jennifer Mendel Whitehill, Jon S Vernick, and Elizabeth M Parker (2012) "Evaluation of Baltimore's Safe Streets Program: effects on attitudes, participants' experiences, and gun violence," *Baltimore, MD: Johns Hopkins Center for the Prevention of Youth Violence*.

Weisburd, David (2015) "The law of crime concentration and the criminology of place," *Criminology*, 53 (2), 133–157.

Westfall, Peter H. and S. Stanley Young (1993) *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*: Wiley-Interscience.

Wheeler, Andrew P, Robert E Worden, and Jasmine R Silver (2019) "The accuracy of the violent offender identification directive tool to predict future gun violence," *Criminal justice and behavior*, 46 (5), 770–788.
Williams, J Patrick (2008) "Emergent themes," *The Sage encyclopedia of qualitative research methods*, 1, 248–249.
Wilson, David B, Leana Allen Bouffard, and Doris L MacKenzie (2005) "A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders," *Criminal Justice and Behavior*, 32 (2), 172–204.
Wilson, Jeremy M and Steven Chermak (2011) "Community-driven violence reduction programs: Examining Pittsburgh's One Vision One Life," *Criminology & Public Policy*, 10 (4), 993–1027.
Wolfgang, Marvin E and Paul E Tracy (1982) *The 1945 and 1958 birth cohorts: A comparison of the prevalence, incidence, and severity of delinquent behavior*: Philadelphia: Center for Studies in Criminology and Criminal Law, University of Pennsylvania.
Wood, George and Andrew V Papachristos (2019) "Reducing gunshot victimization in high-risk social networks through direct and spillover effects," *Nature human behaviour*, 3 (11), 1164–1170.
Ziliak, Steve and Deirdre Nansen McCloskey (2008) *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*: University of Michigan Press.

9 Tables and Figures

Table I: Baseline characteristics

	Control Mean	Treatment Mean	Pairwise p-value
N	1232	1224	
Demographics			
Black	0.969	0.971	0.864
Age	25.3	25.1	0.431
Primary Outcome Components, Counts			
Shooting Victimizations	0.479	0.436	0.119
Shooting & Homicide Arrests	0.079	0.073	0.571
Other Serious Violent-Crime Arrests	0.915	0.878	0.420
Risk Prediction			
Predicted Involvement in a Violent Gun Crime (Risk Score)	0.115	0.114	0.798
Missing Risk Score	0.108	0.080	0.008
Arrest Counts			
All Arrests	17.1	17.7	0.194
Less Serious Violent-Crime Arrests	1.5	1.6	0.960
Drug Crime Arrests	4.9	5.3	0.057
Property Crime Arrests	1.6	1.7	0.643
Other Crime Arrests	8.0	8.3	0.367
Victimization Counts			
All Victimizations	3.5	3.3	0.297
Other Violent Victimizations	2.4	2.3	0.474
Non-Violent Victimizations	0.581	0.560	0.649
Incarceration Measures			
Days Incarcerated	177.3	174.2	0.704
Incarcerated at Randomization	0.042	0.039	0.671
Joint Test			
p-value on F-test			0.302

Notes: Pairwise p-value from test of treatment-control difference using heteroskedasticity-robust standard errors and controlling for randomization strata fixed effects. Arrest and victimization measures include all available CPD data from 1999 (2010 for shooting victimizations) through the time of randomization, with counts winsorized at the top 99^{th} percentile. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. Less serious violent-crime arrests include non-Part I violent-crime arrests, such as simple assault and battery and domestic violence. Other (non-shooting) violent victimizations include sexual assault, robbery, aggravated assault and battery, and simple assault and battery. Non-violent victimizations include all other incidents such as burglary, stalking, and threats. Risk score is missing for 231 individuals who did not have at least one arrest or two victimizations within the 50 months prior to randomization (risk score N for control = 1,099 and for treatment = 1,126). Race is missing for 38 individuals (race N for control = 1,211 and for treatment = 1,207). Joint test includes randomization strata fixed effects and the covariates listed here, excluding all arrests and all victimizations since they are linear combinations of other variables.

Table II: Baseline characteristics and realized risk by pathway

	All	Algorithm	Outreach	Re-entry	P-value, Test of Pathway Difference
Baseline					
N	2456	1232	878	346	
Age	25.2	24.6	25.6	26.4	<.001
Black	0.970	0.963	0.986	0.953	<.001
Shooting Victimizations	0.458	0.631	0.287	0.275	<.001
Shooting & Homicide Arrests	0.076	0.080	0.064	0.092	0.179
Other Serious Violent-Crime Arrests	0.897	1.0	0.667	1.0	<.001
Ever Shot	0.347	0.463	0.233	0.225	<.001
Ever Arrested	0.977	1.0	0.945	0.974	<.001
All Arrests	17.4	20.4	13.6	16.1	<.001
Ever Victimized	0.835	0.908	0.768	0.743	<.001
All Victimizations	3.4	4.5	2.3	2.3	<.001
Days Incarcerated	175.7	129.3	155.3	392.8	<.001
Predicted Involvement in a Violent Gun Crime	0.114	0.137	0.089	0.080	<.001
Realized Risk Among Controls					
N	1232	616	438	178	
Shooting & Homicide Victimizations	0.110	0.106	0.114	0.118	0.874
Shooting & Homicide Arrests	0.027	0.023	0.032	0.028	0.683
Other Serious Violent-Crime Arrests	0.054	0.057	0.050	0.056	0.912
Ever Shot or Killed	0.103	0.106	0.100	0.101	0.961
Ever Arrested	0.636	0.700	0.578	0.562	<.001
All Arrests	1.7	1.9	1.4	1.3	<.001
Ever Victimized	0.315	0.354	0.285	0.253	0.009
All Victimizations	0.526	0.670	0.397	0.343	<.001
Days Incarcerated	79.3	88.4	62.7	88.7	0.009
Involved in a Violent Gun Crime	0.159	0.174	0.148	0.135	0.335

Notes: Top panel shows baseline characteristics. Arrest and victimization measures include all available CPD data from 1999 (2010 for shooting victimizations) through the time of randomization, with counts winsorized at the top 99^{th} percentile. "Predicted Involvement in a Violent Gun Crime" shows the risk score: the predicted probability of being a victim or an arrestee in a violent crime involving a gun during the next 18 months. "Involved in a Violent Gun Crime" shows the actual realized rate of that same outcome over the 18 months after randomization for the control group. With the exception of that 18-month outcome period, the rest of the bottom panel shows control means during the 20-month outcome period. P-value from joint test of the null that referral pathway means are equal using heteroskedasticity-robust standard errors. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. N for the risk score is 2,225 for All, 1,232 for algorithm referrals, 717 for outreach referrals, and 276 for re-entry referrals.

Table III: Program take-up rates and participation conditional on take-up, by pathway

		Conditional on Take-up						
		Work Hours				CBT/Training Hours	Total Hours	Total Earnings
	Take-up Rate	25th Percentile	50th Percentile	75th Percentile	Average	Average	Average	Average
All Participants	55%	149	437	868	560	159	719	$9,651
Algorithm	37%	91	388	830	517	142	659	$8,838
Outreach	78%	196	502	942	615	172	787	$10,465
Re-entry	60%	128	438	792	506	157	663	$9,183

Note: Take-up defined as attending the first day of READI orientation. Hours and earnings are averages over the entire 20-month outcome period (including after the start of COVID-19) among those who took up. The maximum number of hours someone could have participated during the 20-month outcome period depends on their time from randomization to take-up; for the pre-COVID period, see Appendix Table A.II. Work hours correspond to time spent at a worksite. CBT/training hours correspond to time spent in group CBT sessions, professional development sessions, and online trainings. Hours and earnings are limited to men who took up and appear in the payroll data, which excludes 124 men who either took up prior to consent forms allowing the release of their payroll data being distributed or who attended orientation but did not start work within 20 months of randomization. Participation records for CBT and training before April 2020 are incomplete; hours shown are extrapolated based on available data. For additional details on hours and earnings data, see Appendix A.2.4.

Table IV: READI's estimated effects on serious violence involvement

	Estimates				P-values		
	CM	ITT	CCM	TOT	Observed ITT	FWER	FDR-q
Primary Index of Serious Violence							
	0	-0.0266	0.0155	-0.0487	0.257		
		(0.0234)		(0.0413)			
Primary Outcome Components, Counts							
Shooting & Homicide Victimizations	0.1104	-0.0072	0.1105	-0.0132	0.593	0.833	0.733
		(0.0134)		(0.0237)			
Shooting & Homicide Arrests	0.0268	-0.0120	0.0340	-0.0220	0.045	0.126	0.134
		(0.0060)		(0.0106)			
Other Serious Violent-Crime Arrests	0.0544	0.0034	0.0551	0.0063	0.733	0.728	0.733
		(0.0101)		(0.0177)			

Notes: N = 2,456. Primary index standardizes each of the three components shown in the bottom panel based on the control group's distribution and takes their unweighted average. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. Multiple hypothesis testing adjustments define the three components of the primary index as a family. FWER p-values control for the family-wise error rate using the Westfall and Young (1993) method for step-down resampling. FDR q-values control for the false discovery rate using the Benjamini and Hochberg (1995) method. CM shows control means; ITT shows intent-to-treat estimates; CCM shows control complier means; and TOT shows treatment-on-the-treated estimates from an IV regression. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses.

Table V: READI's estimated effects on serious violence involvement, by pathway

	Estimates				P-values		
	CM	ITT	CCM	TOT	Observed ITT	FWER	FDR-q
Primary Index of Serious Violence by Pathway							
Algorithm (N = 1232)	-0.0097	0.0360	-0.0640	0.0990	0.298	0.529	0.312
		(0.0346)		(0.0917)			
Outreach (N = 878)	0.0085	-0.1007	0.0563	-0.1293	0.008	0.029	0.024
		(0.0379)		(0.0469)			
Re-entry (N = 346)	0.0126	-0.0600	0.0501	-0.1003	0.312	0.529	0.312
		(0.0594)		(0.0958)			
Primary Outcome Components by Pathway, Counts							
Algorithm							
Shooting & Homicide Victimizations	0.1055	0.0231	0.0654	0.0635	0.243	0.552	0.415
		(0.0197)		(0.0524)			
Shooting & Homicide Arrests	0.0227	-0.0048	0.0222	-0.0133	0.572	0.571	0.572
		(0.0085)		(0.0226)			
Other Serious Violent-Crime Arrests	0.0568	0.0169	0.0468	0.0465	0.277	0.552	0.415
		(0.0155)		(0.0411)			
Outreach							
Shooting & Homicide Victimizations	0.1142	-0.0438	0.1320	-0.0562	0.040	0.080	0.061
		(0.0214)		(0.0264)			
Shooting & Homicide Arrests	0.0320	-0.0255	0.0415	-0.0328	0.009	0.025	0.028
		(0.0098)		(0.0121)			
Other Serious Violent-Crime Arrests	0.0502	-0.0045	0.0583	-0.0059	0.783	0.792	0.783
		(0.0165)		(0.0203)			
Re-entry							
Shooting & Homicide Victimizations	0.1180	-0.0214	0.1358	-0.0358	0.546	0.823	0.819
		(0.0355)		(0.0569)			
Shooting & Homicide Arrests	0.0281	-0.0028	0.0346	-0.0046	0.869	0.883	0.869
		(0.0171)		(0.0274)			
Other Serious Violent-Crime Arrests	0.0562	-0.0246	0.0612	-0.0412	0.204	0.562	0.613
		(0.0194)		(0.0312)			

Notes: N = 2,456. Estimates for each outcome are from a single regression that interacts pathway indicators with treatment. Primary index standardizes each of the three components shown in the bottom panel based on the control group's distribution and takes their unweighted average. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. For multiple hypothesis testing adjustments within the primary index, we define the three different pathways as a family. For the component adjustments, we define the three different outcomes within each pathway as a family. FWER p-values control for the family-wise error rate using the Westfall and Young (1993) method for step-down resampling. FDR q-values control for the false discovery rate using the Benjamini and Hochberg (1995) method. CM shows control means; ITT shows intent-to-treat estimates; CCM shows control complier means; and TOT shows treatment-on-the-treated estimates from an IV regression. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses. F-tests of the null hypothesis that treatment effects are equal across the three pathways are as follows: Primary Index: $p = 0.028$; Shooting & Homicide Victimizations: $p = 0.069$; Shootings & Homicide Arrests: $p = 0.236$; Other Serious Violent-Crime Arrests: $p = 0.235$.

Table VI: READI's estimated effects on serious violence involvement, by baseline risk level

	Estimates				P-values		
	CM	ITT	CCM	TOT	Observed ITT	FWER	FDR-q
Primary Index of Serious Violence by Risk Level							
Over Median (N = 1112)	0.0850	-0.0669	0.1717	-0.1482	0.099	0.259	0.298
		(0.0405)		(0.0866)			
Under Median (N = 1113)	-0.0511	0.0100	-0.0671	0.0169	0.754	0.942	0.852
		(0.0321)		(0.0510)			
Missing (N = 231)	-0.1489	-0.0092	-0.1482	-0.0128	0.852	0.942	0.852
		(0.0492)		(0.0697)			
Primary Outcome Components by Risk Level, Counts							
Over Median							
Shooting & Homicide Victimizations	0.1434	-0.0193	0.1573	-0.0427	0.388	0.619	0.582
		(0.0223)		(0.0475)			
Shooting & Homicide Arrests	0.0376	-0.0206	0.0575	-0.0456	0.045	0.125	0.135
		(0.0103)		(0.0219)			
Other Serious Violent-Crime Arrests	0.0771	-0.0048	0.1017	-0.0108	0.784	0.781	0.784
		(0.0175)		(0.0373)			
Under Median							
Shooting & Homicide Victimizations	0.0906	0.0056	0.0853	0.0093	0.770	0.773	0.770
		(0.0191)		(0.0303)			
Shooting & Homicide Arrests	0.0203	-0.0069	0.0227	-0.0112	0.410	0.691	0.615
		(0.0083)		(0.0132)			
Other Serious Violent-Crime Arrests	0.0407	0.0135	0.0292	0.0224	0.324	0.691	0.615
		(0.0137)		(0.0218)			
Missing							
Shooting & Homicide Victimizations	0.0526	-0.0109	0.0614	-0.0159	0.754	0.971	0.754
		(0.0348)		(0.0493)			
Shooting & Homicide Arrests	0.0075	0.0048	0.0076	0.0075	0.715	0.971	0.754
		(0.0133)		(0.0188)			
Other Serious Violent-Crime Arrests	0.0150	-0.0060	0.0090	-0.0090	0.698	0.971	0.754
		(0.0154)		(0.0218)			

Notes: N = 2,456. Estimates for each outcome are from a single regression that interacts indicators for above/below-median risk score and missing risk score with treatment. The median risk score in the study sample, the predicted probability at baseline of being arrested for or the victim of a violent gun crime in the next 18 months, is 0.11. Primary index standardizes each of the three components shown in the bottom panel based on the control group's distribution and takes their unweighted average. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. For multiple hypothesis testing adjustments within the primary index, we define the three different risk score groups as a family. For the component adjustments, we define the three different outcomes within each risk score group as a family. FWER p-values control for the family-wise error rate using the Westfall and Young (1993) method for step-down resampling. FDR-q values control for the false discovery rate using the Benjamini and Hochberg (1995) method. CM shows control means; ITT shows intent-to-treat estimates; CCM shows control complier means; and TOT shows treatment-on-the-treated estimates from an IV regression. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses. F-tests of the null hypothesis that treatment effects are equal across the three risk score groups are as follows: Primary Index: $p = 0.330$; Shooting & Homicide Victimizations: $p = 0.700$; Shootings & Homicide Arrests: $p = 0.299$; Other Serious Violent-Crime Arrests: $p = 0.569$

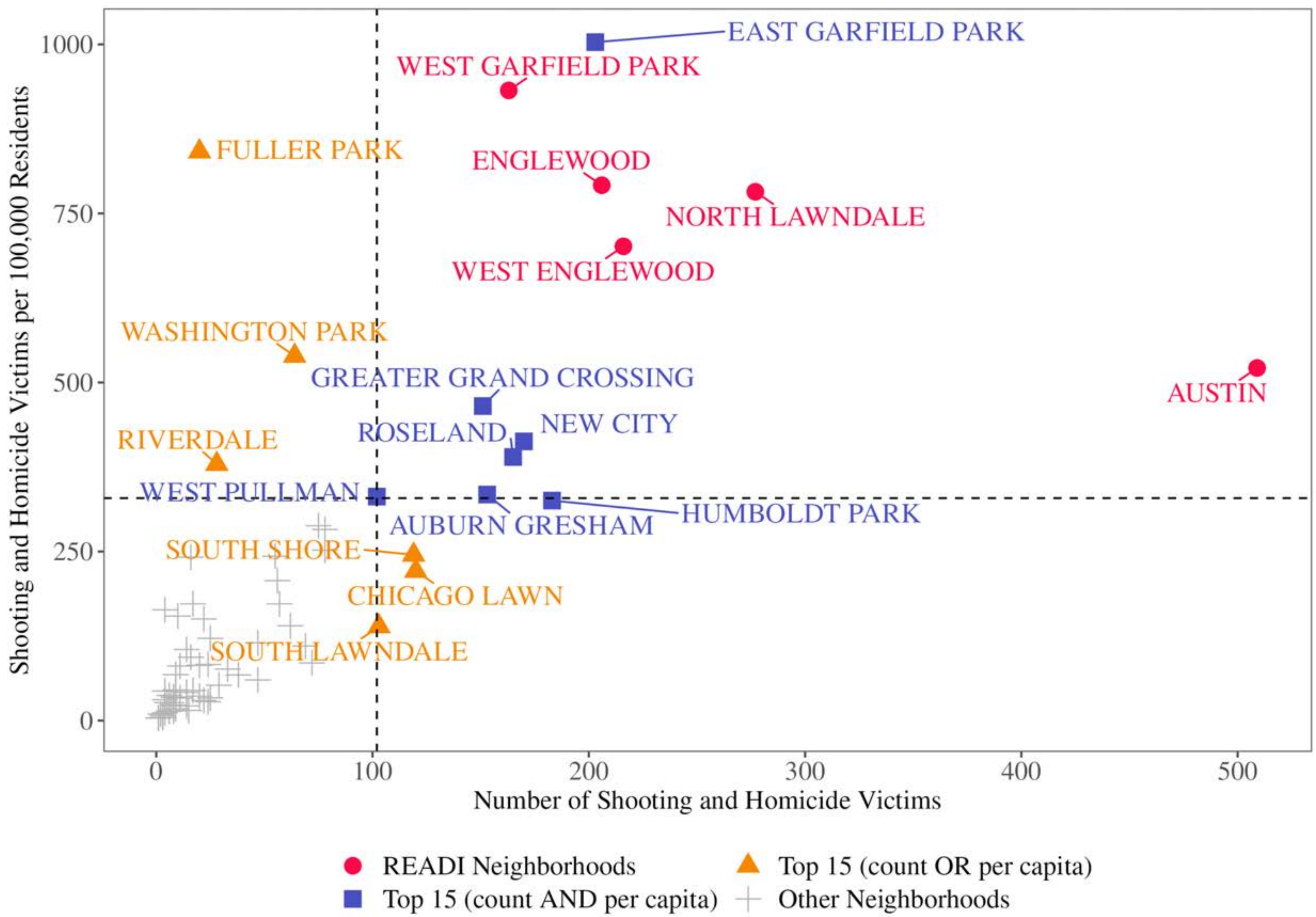
Table VII: READI's estimated effect on the social costs of crime

	Less Inclusive Estimates				More Inclusive Estimates			
	CM	ITT	CCM	TOT	CM	ITT	CCM	TOT
Social Cost of Victimization								
READI Sample Victims	\$147,173	-\$22,741	\$149,680	-\$41,724	\$390,251	-\$56,719	\$400,670	-\$104,064
		(\$33,363)		(\$58,848)		(\$82,482)		(\$145,494)
READI Sample Offenders	\$127,058	-\$69,594**	\$188,796	-\$127,686**	\$1,061,717	-\$435,114**	\$1,442,056	-\$798,317**
		(\$29,140)		(\$51,542)		(\$201,407)		(\$356,080)
Social Cost of Punishment								
Legal System Costs	\$10,776	-\$4,535**	\$14,819	-\$8,320**	\$14,679	-\$5,113**	\$18,832	-\$9,381**
		(\$1,916)		(\$3,389)		(\$2,065)		(\$3,652)
Productivity Loss from Incarceration	\$4,506	-\$2,106**	\$6,369	-\$3,864**	\$6,315	-\$2,376**	\$8,232	-\$4,359**
		(\$894)		(\$1,582)		(\$964)		(\$1,705)
Total Social Cost of Crime	\$289,513	-\$98,976**	\$359,665	-\$181,595**	\$1,472,962	-\$499,321**	\$1,869,790	-\$916,120**
		(\$45,791)		(\$80,911)		(\$219,071)		(\$387,313)
Social Cost of Program								
Administrative Costs		\$33,510		\$61,483		\$33,510		\$61,483
Transfer to Participants		-\$5,260		-\$9,651		-\$5,260		-\$9,651
Net READI Costs		\$28,250		\$51,832		\$28,250		\$51,832
Benefit-Cost Ratio		3.5:1		3.5:1		17.7:1		17.7:1

Notes: N = 2,456. Outcome variables in top panel are different measures of social costs from crime. Negative point estimates show gains from reduced crime. Crime cost estimates, from Cohen and Piquero (2009) and inflated to 2017 dollars, aim to quantify harm to victims from each crime, the productivity loss to offenders from their involvement in the legal system, and the cost to the government of running the legal system. The less inclusive estimates use bottom-up crime cost estimates and only assign costs to observed arrests and victimizations (ignoring both offending that does not result in arrest and arrests of non-study individuals who victimized READI participants). The more inclusive estimates use willingness-to-pay crime cost estimates and extrapolate social harm based on estimates of clearance rates (i.e., for each crime type's clearance rate, C, we assume there are $1/C$ crimes committed for each observed arrest and C arrests made for each observed victimization), as well as victimization reporting rates. Program costs calculated from spending reported by Heartland Alliance over the 20-month outcome period. CM shows control means; ITT shows intent-to-treat estimates; CCM shows control complier means; and TOT shows treatment-on-the-treated estimates from an IV regression. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Figure I

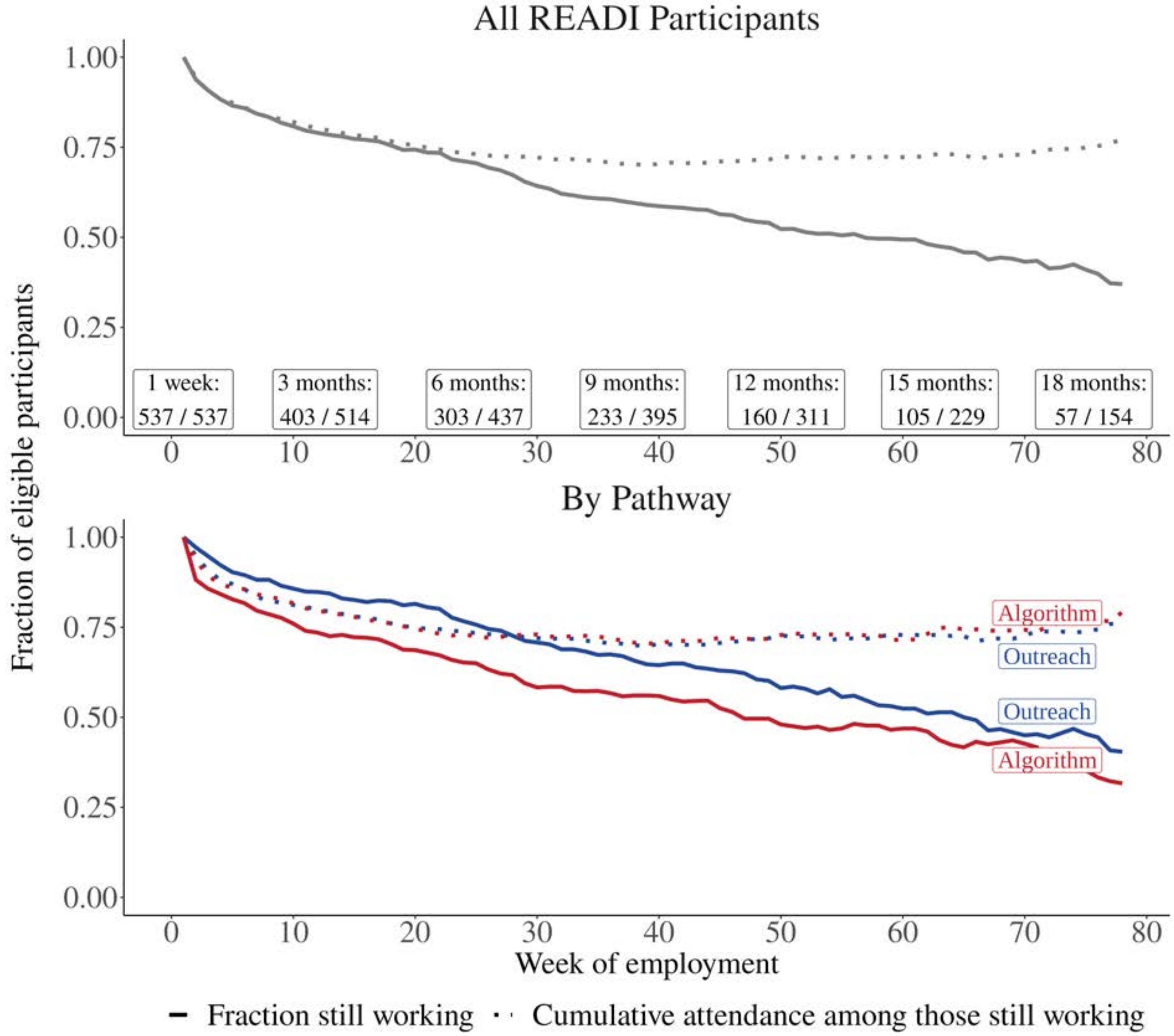
Shooting victims per 100,000 residents (2016), by neighborhood



Notes: Plot shows counts and rates of shooting and homicide victims in each of Chicago's 77 neighborhoods in 2016. Dashed lines represent top 15 neighborhoods for each dimension.

Figure II

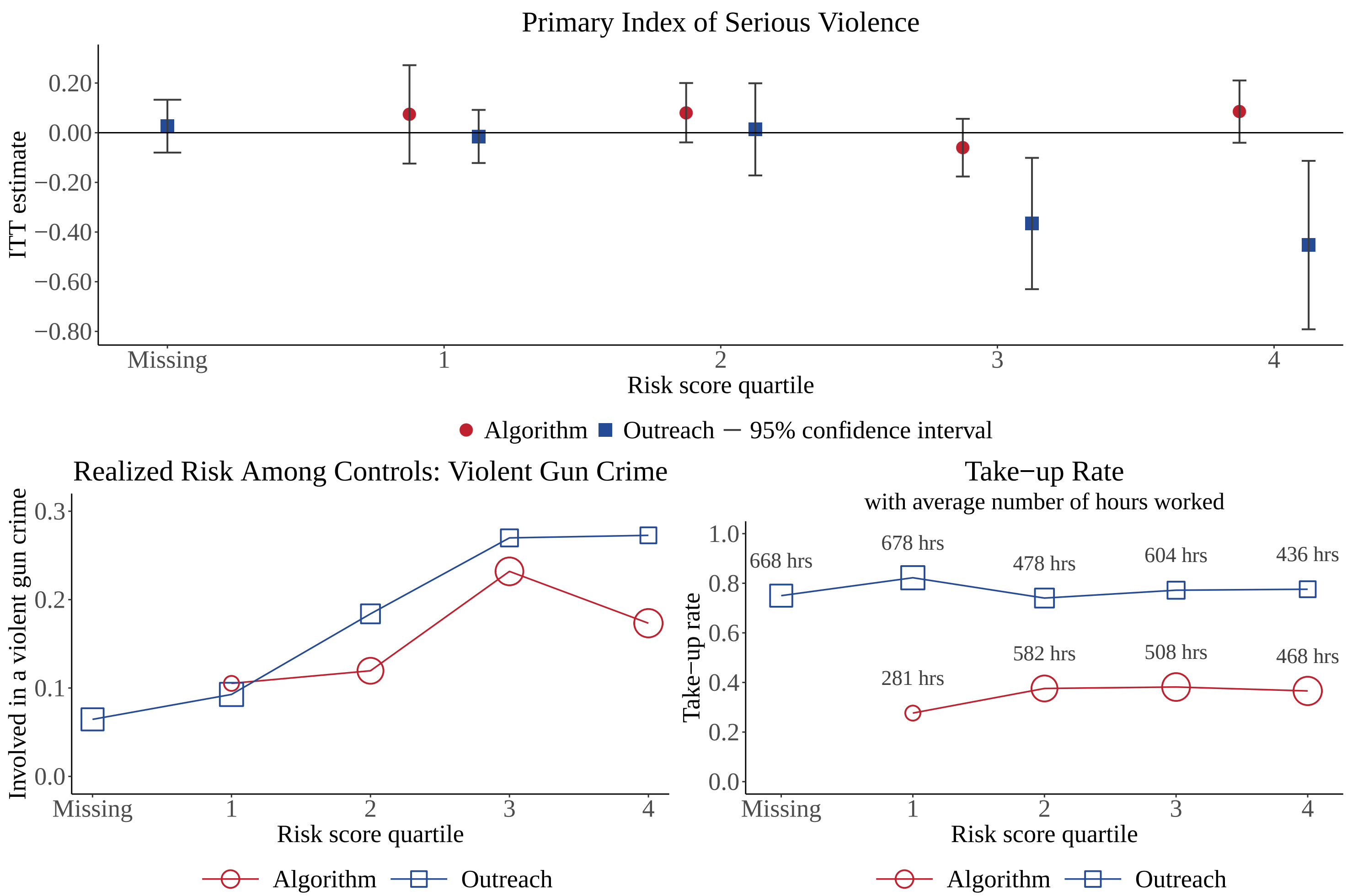
READI job retention, overall and by pathway



Notes: Figure shows two measures of job retention for men who started READI measured from payroll data. The solid line shows the proportion of participants who work at least once after the time shown on the x-axis conditional on observing them for that long. The boxes show the number of workers contributing to each point. The dotted line shows the average proportion of possible weeks worked among those still working at each point in time. At 18 months after first taking up, N = 19/60 algorithm referrals and N = 38/94 outreach referrals are still observed working. Since COVID-19 changed both what "participation" looked like in practice as well as payroll policies, we report these measures of retention using only data through the start of the pandemic. Because there were too few re-entry participants with sufficient pre-COVID data to measure retention, they are omitted. For a description of the pandemic's impact on READI, see Section 2.3 and Appendix A.5.3. For retention measures inclusive of the pandemic period, see Appendix Figure A.III.

Figure III

READI estimated effects, realized risk, and take-up by pathway and predicted risk



Notes: The top panel shows coefficient estimates and 95 percent confidence intervals (using heteroskedasticity-robust standard errors) on three-way interactions of pathway indicators, risk score quartile indicators, and an indicator for being randomized to receive a READI offer, from a regression of the primary index on baseline covariates, randomization strata fixed effects, and all two-way interactions of pathway indicators, risk score quartile indicators, and an indicator for being randomized to receive a READI offer. Bottom left panel shows the realized rate of involvement in a violent gun crime as a victim or an arrestee during the 18 months after randomization by quartile of the risk score, which is the predicted probability at baseline of the same outcome, separately for algorithm and outreach referrals. Bottom right panel shows the take-up rate, defined as the share of the treatment group attending the first day of READI orientation, by quartile of the risk score, separately for algorithm and outreach referrals. Markers in both of the bottom panels are weighted to reflect the share of algorithm and outreach referrals in each risk score quartile. N = 231 for the missing risk score group (of whom 161 are outreach referrals). For ITT estimates of each individual index component by pathway and risk score quartile, see Appendix Figure A.VI.

Appendix

A.1 Research on policing and CVIs

The concentrated, place-based nature of much violent crime is well documented (e.g., Weisburd, 2015; Abt, 2019; Blattman et al., 2021), and targeted interventions for reducing concentrated gun violence have a long history. For instance, "hot spots" and "proactive" policing models focus resources on small geographic areas with a high risk of gun crime occurring. "Focused deterrence" and other related strategies apply sanctions and provide services to specific people and groups at high risk of involvement in future gun violence, often with community involvement. Both approaches show some success in reducing violence overall and gun violence specifically (e.g., Grogger, 2002; Sherman and Rogan, 1995; Braga et al., 2001, 2014, 2018; Ridgeway et al., 2019; Wood and Papachristos, 2019). But increased police attention can impose significant costs on its targets (Ang, 2021; Geller et al., 2014; Jones, 2014; Chalfin et al., 2022). It can also cause harassment of or harm to residents of hot spots who are not engaged in any criminal activity but are nonetheless subjected to greater police scrutiny. This concern has renewed interest in community violence interventions (CVIs)—programs such as READI that offer alternatives to law enforcement-focused approaches.

Among the most widely studied and adopted CVIs focused on shootings is violence interruption, exemplified by organizations such as Cure Violence.[1a] Violence interruption relies on outreach workers to mediate active disputes, foster community norms of non-violence, and refer the highest-risk individuals to available services (Butts et al., 2015a). In addition to community-wide education and mobilization efforts, neighborhood-based outreach workers identify and leap into imminent or active disputes, especially an initial shooting, and try to avert cycles of retaliatory violence by helping involved parties cool emotions, step back from immediate conflict, and find non-violent ways to respond, save face, or achieve justice. There are parallels to the emotional regulation and slower thinking encouraged by CBT-informed programs such as READI. By its nature, however, mediation and other violence interruption techniques tend to be more immediate and reactive. This interruption approach and the READI model are distinct and, in principle, complementary—akin to emergency and preventive medicine. Reviewing the literature on interventions adopting the Cure Violence model, Butts et al. (2015a) find the evidence of their effectiveness to be "mixed at best." Some studies report large reductions in homicides and shootings in specific neighborhoods of Chicago and Baltimore that researchers attribute to the intervention (Skogan et al., 2008; Webster et al., 2012), while others find no such impacts or even evidence of adverse effects (Wilson and Chermak, 2011). A recent study by Buggs et al. (2022) tries to overcome

[1a] See, e.g., Skogan et al. (2008); Wilson and Chermak (2011); Butts et al. (2015b); Fox et al. (2015); Webster et al. (2012); Picard-Fritsche and Cerniglia (2013); Buggs et al. (2022). For an overview, see Butts et al. (2015a).

the confounding challenges in this literature by using synthetic control methods to construct more reliable comparison neighborhoods. Overall, neighborhood-level interventions are quite challenging to evaluate given the difficulty of finding comparable comparison groups when many determinants of violence are unobserved to the researcher.

The READI model is closer to a relatively new set of CVIs that deliver intensive preventative services to people thought to be at high risk of shooting involvement. These programs—which include Advance Peace, Roca, Turn90, CRED, GRID, and READI—all rely on local outreach workers to find and engage clients, usually young men of color, with extensive prior criminal legal system involvement and exposure to violence and trauma. Although they differ somewhat in the composition and duration of the services they provide, most offer some combination of financial assistance or supported employment, as well as counseling or CBT. Advance Peace, operating in high-violence neighborhoods within several California cities, provides clients with 18 months of mentorship, life skills coaching, and financial support (Corburn and Fukutome, 2021). Roca, operating in several Massachusetts sites and Baltimore, Maryland, provides clients intensive case management, supported education or employment opportunities, and CBT, with an average engagement of approximately two years (Abt Associates, 2021). Turn90 (formerly Turning Leaf), operating in Charleston, South Carolina, provides men returning from prison with 150 hours of CBT, case management, and a transitional job. CRED, operating in several Chicago neighborhoods, provides clients with life coaching and counseling, academic support, and a transitional job. GRID, or the Gang Reduction Initiative of Denver, develops case plans and provides service coordination for youth and adults looking to exit gang involvement.

These newer CVIs are promising and represent potentially cost-effective models, but with one exception, there is no causal evidence of their effectiveness so far. An evaluation of Advance Peace follows the approach taken by studies of violence interruption programs and measures its impact at the city- and neighborhood-level, but lacks a comparison group (Corburn et al., 2022). An implementation evaluation of Roca that similarly lacks a comparison group reports the relationship between participants' outcomes and the intensity of their engagement with programming (Abt Associates, 2021), though our understanding is that a separate impact evaluation remains underway. A report summarizing preliminary evidence on CRED finds that participants experience 50 percent fewer shooting victimizations in the 18 months after starting programming than they did before, but a quasi-experimental comparison with similar young men not in CRED does not find statistically significant effects.[2a] The one notable exception is GRID, which was the subject of a randomized trial that found

[2a] https://www.ipr.northwestern.edu/documents/reports/ipr-n3-rapid-research-reports-cred-impact-aug-25-2021.pdf

participants were more likely to identify as gang members, less likely to self-report having recently committed a violent offense, but no less likely to be arrested, charged, convicted, or incarcerated for any kind of criminal offense (Pyrooz et al., 2023). Due to the limited available causal evidence on this type of targeted CVI programming, its impact remains an open question.

A.2 Data details

A.2.1 Chicago Police Department

We use data from the Chicago Police Department (CPD) to develop most of our baseline covariates, in addition to our main outcome measures of serious violence involvement. The six different CPD datasets we use are: arrests from 1999 to present (including juvenile arrest information), general victimizations reported to CPD from 1999 to present, reported crimes from 1999 to present (which are not tied to individual victims), shooting victimizations from 2010 to present, homicide victimizations from 1999 to present, and homicide offenders from 1999 to present.[3a] Our measures of age and race/ethnicity also come from CPD records, with the latter likely to be measured with error since it captures officer impressions rather than self-identification.

A critical prerequisite to computing baseline and outcome measures is having the ability to link READI study members to their records across CPD datasets. Only the arrests data contain a unique person identifier based on fingerprint (IR number) that allows individuals to be tracked over time. As a result, we must use a probabilistic record linkage algorithm to associate each unique study member with their records across different CPD datasets. For details on the algorithm itself, Name Match, see McNeill and Jelveh (2022). We describe the basics of the linking procedure below.

The high-level goal of the record linkage algorithm is to divide the records found in six datasets—the READI study member roster and CPD arrests, general victimizations, shooting victimizations, homicide victimizations, and homicide offenders—into clusters, or groups of records that refer to the same person. To do that, we first train a probabilistic matching algorithm to learn how identifying information in two arrest records differs when the records do and do not belong to the same person; then use that algorithm to predict the likelihood that pairs of records across all six data sources belong to the same person; and finally group together records with a high probability of belonging to the same person.

The algorithm consists of three stages: learning, predicting, and aggregating. In the

[3a] There is imperfect overlap between the arrests and homicide offenders datasets. One possible reason for this discrepancy is if the homicide offenders dataset contains information about individuals whom CPD suspects of being homicide offenders but not all of whom have been arrested. Alternatively, this discrepancy may be an artifact of how these data are received as snapshots from CPD.

learning stage, a supervised learning model is trained to identify pairs of arrest records that refer to the same person using information on distances between the records' identifying fields. Specifically, we first compute distance metrics like edit distance, indicators for whether two fields match exactly or sound the same (i.e., Soundex), and days between incidents using the following fields: first name, last name, date of birth (DOB), age, middle initial, race, gender, home address, and date of CPD incident. We then leverage the fingerprint-based IR number in the arrest data to generate ground-truth labels on match status for pairs of arrest records. Finally, this dataset of distance between identifying fields and known match status is then used to train a random forest classifier capable of estimating P(match) for record pairs where match status is unknown. Notice that this algorithm will learn that, e.g., if addresses vary a lot across arrest records belonging to the same person (i.e., belonging to the same IR number), then address information is assigned a lower weight in predicting whether two records belong to the same person.

In the predicting stage, we estimate P(match) for pairs of records across the six input datasets. That is, we obtain predicted match status between each READI study member's identifying information—first name, last name, DOB, age, gender—and CPD records, as well as between CPD records and other CPD records. By considering links between pairs of CPD records and not just direct links between READI study roster records and CPD records, we reduce the risk of false negative links due to records with typos in multiple identifying fields.

In the final aggregating stage, predicted links are chained together into groups of records that are assigned the same new unique person identifier. This person identifier can then be used to identify each READI study member's various CPD records. During aggregation, the algorithm follows researcher-specified constraints that are customized to fit the context of the data. For our linkage, we specify the following constraints. First, all records with the same IR number from 2010 onward are linked regardless of their estimated P(match), and a cluster can have at most one such IR number. Because IR number is based on an individual's fingerprint, two records with the same IR number are known to refer to the same person, just as two predicted links with different IR numbers are known to be false positives. We explicitly enforce these known matches and prevent these false positives in cases where the IR numbers were from 2010 onward, when the this field is known to be more reliable. Note that study members can still be linked to any record without an IR number and to arrest records from before 2010; we just do not have enough confidence in the pre-2010 IR numbers to use them to enforce hard constraints.

Second, a homicide victimization record cannot link to another record if the homicide record's event date is before the other record's event date. Third, some victimization records do not have DOB information. To reduce the chance of this causing false positive links, we introduce a constraint that if at least one record in a pair is missing DOB information, then the ages on the records must be within 3 years of each other. We also enforce that if at least one record in a potential cluster is missing exact DOB information, all other records in the cluster not missing DOB information must have similar DOBs.

It is important to our research design to ensure that links formed between a study member and their CPD records cannot be affected by their treatment status. We accomplish this with a two-stage matching process that modifies the linking algorithm in several key ways. During the learning stage, we limit the training data to only include arrest records from before the first study member's randomization date. When aggregating predicted links, we evaluate links in chronological order and enforce one final constraint: once a READI study member is linked to a CPD record, the study member cannot not be linked to any more CPD records from before their randomization date. This process makes it impossible for post-randomization police involvement, which could be shaped by READI itself, to affect which pre-randomization CPD records are linked to a study member.

The table below details the share of records with missingness in each identifying field, separately by dataset and time period (pre- or post-randomization). The matching algorithm relies most heavily on first name, last name, and DOB to estimate high quality P(match). Fortunately, there is no missingness in first name or last name, and the missingness in DOB is highly concentrated in non-shooting victimizations from well before the study began. Links between study members and their arrest records are likely to be the most well measured, due to no DOB missingness and the extra constraints we enforce using IR number. Links between study members and shooting victimizations are also well measured due to nearly complete DOB information in the shooting victimizations dataset. Of our main outcomes, the homicide victimizations are the most difficult to link due to the lack of DOB information. However, 80 percent of the homicide victimization records have a listed IR number and therefore are automatically linked to arrest records. The constraints described above are designed to help prevent false positives in the 20 percent of homicide victimizations with neither a DOB nor an IR number.

CPD data missingness

		Arrests	Victimizations	Shooting Victimizations	Homicide Victimizations
Pre-Randomization	IR Number	0.1% missing (Among post-2010 IRs that are used)	100% missing	100% missing	20% missing
	DOB	–	68% missing (Missingness decreases steadily from 98% in 1999 to 25% in 2017)	1% missing	100% missing
	Age	–	4% missing	–	–
	Middle Initial	50% missing	74% missing	100% missing	72% missing
	Race	–	5% missing	1% missing	–
	Gender	–	3% missing	–	–
	Home Address	1% missing	1% missing	–	100% missing
Post-randomization	IR Number	2% missing	100% missing	100% missing	20% missing
	DOB	–	18% missing	1% missing	100% missing
	Age	–	6% missing	–	–
	Middle Initial	36% missing	64% missing	100% missing	51% missing
	Race	–	5% missing	1% missing	–
	Gender	–	2% missing	–	–
	Home Address	1% missing	1% missing	–	100% missing

Notes: In addition to the identifying fields listed above, all CPD datasets have first and last name with no missingness. The READI study members roster file contains first name, last name, DOB, age, gender (all men), and randomization date. The green cells highlight the information used during the linking algorithm's learning stage. A fifth CPD dataset containing homicide offender records was included in the record linkage process, but was ultimately not used for covariate or outcome definition. For simplicity, this dataset is excluded from the table above.

Given the use of IR numbers from 2010 onward as ground truth, anyone linked to a pre-randomization IR number will necessarily be linked to all post-randomization arrests associated with that IR number (in addition to all other records that match on other identifying information). In practice, 96 percent of the sample matched to a pre-randomization IR number, with a slight treatment-control difference in who is matched to a pre-randomization IR number. When we regress an indicator for having matched to a pre-randomization IR number on a treatment indicator and strata fixed effects, the treatment group is slightly more likely to have matched $\beta = 0.017$ relative to a control mean of 0.956 ($p = 0.02$). This difference is driven mostly by the re-entry pathway (outreach: $\beta = 0.027$ from a baseline of .911, $p = 0.12$; re-entry: $\beta = 0.051$ from a baseline of .916, $p = 0.05$). Algorithm referrals are never missing a pre-randomization IR number. One concern with this imbalance is that individuals without pre-randomization IRs may be less likely to link to post-period arrests. However, the scale of the baseline IR missingness is small; IR is not the only information we use to match outcome data to study members; and the directionality of the imbalance suggests that, if anything, the treatment effects we estimate for arrest outcomes are slightly conservative because treatment group members would be slightly more likely to be linked to their post-period CPD records.

A.2.2 CPD outcome definitions

The study's primary outcome is an index comprising three components: (1) shooting and homicide victimizations, (2) shooting and homicide arrests, and (3) other serious violent-crime arrests. We define each of these measures using FBI codes and other fields in records within four CPD datasets: arrests, general victimizations, shooting victimizations, and homicide victimizations.[4a]

Shooting and homicide victimizations CPD stores information about victimizations in both a general dataset and in datasets specific to homicides and non-fatal shootings. To be labeled a homicide victimization, an individual must appear as the victim in either a homicide victimization record or a general victimization record with FBI code 01A (first and second degree murder) or 01B (involuntary manslaughter).

To be labeled a (non-fatal) shooting victimization, an individual must appear as the victim in a shooting victimization record. We do not rely on the general victimizations dataset to identify shooting victimizations because it contains insufficient information to do so reliably. For example, many shootings are classified as aggravated batteries in the general victimizations file; however, incidents involving a knife injury are also often classified as aggravated batteries. In addition, the general victimizations file may categorize shooting incidents related to armed robbery simply as robberies, with no mention of a gun or gun discharge.

Shooting and homicide arrests To be labeled a homicide arrest, an individual must appear as the offender in an arrest record with FBI code 01A or 01B as the top charge.

[4a] `https://gis.chicagopolice.org/pages/crime_details`

Although CPD stores information on secondary charges associated with an arrest, they have advised us that this information is less reliable and to focus instead on the top charge. The only exception we make to this concerns instances where someone is arrested pursuant to a warrant; in these cases, the FBI code for the top charge is WRT, and we look for a supporting charge with FBI code 01A or 01B to label this a homicide arrest.

To be labeled a shooting arrest, it is necessary but insufficient for an individual to appear as the offender in an arrest record with FBI code 04A (aggravated assault) or 04B (aggravated battery) as the top charge. To isolate shooting arrests from other aggravated assault or battery arrests, we rely on two additional types of information: statute descriptions and shooting victimization records. For an aggravated assault (04A) arrest, the statute description must include the word "DISCH" (discharge) to be labeled a shooting arrest. For an aggravated battery (04B) arrest, we require either that the statute description include "GUN" or "FIREARM" but not "NO FIREARM", or that there exist a shooting victimization record associated with the case, to be labeled a shooting arrest. Our use of the shooting victimization records is intended to capture cases where the statute description is insufficiently clear about whether a firearm was used but where the existence of a shooting victim strongly suggests that one was; in practice, only one shooting arrest in the outcome period for our sample turns on this fact, and the statute description in that case is "AGG BATTERY/USE DEADLY WEAPON". Finally, like with homicide arrests, if an arrest has an FBI code for the top charge of WRT, the supporting charge is either 04A or 04B, and the other criteria outlined in this paragraph are met, then it is labeled a shooting arrest.

Other serious violent-crime arrests To be labeled an other serious violent-crime arrest, an individual must appear as the offender in an arrest record that does not meet the definition of either a homicide or shooting arrest described above, and that has one of the FBI codes for a violent Part I offense as its top charge: sexual assault (02), robbery (03), aggravated assault (04A), or aggravated battery (04B).

Finally, arrests are included in our 20-month outcomes only if the underlying incident (rather than the arrest itself) occurred during that period.

Many studies do not separate out non-fatal shootings from other aggravated assaults and batteries, partly because few are powered to detect changes in the rarer shooting incidents. As such, our definition necessarily deviates from standard FBI classification codes. Because the study sample was selected for READI in a particular way, it is difficult to use rates of shooting incidents from other studies of violence or gangs as a benchmark for assessing whether our shooting definition is plausibly capturing the right set of incidents. Nonetheless, we note several ways in which our classification of shootings is consistent with outside data on potentially lethal violence.

On the arrest side, Kapustin et al. (2017) show that in the years leading up to READI, Chicago's clearance rates for non-fatal shootings ranged from 5 to 11 percent. If those rates carried over into our outcome years, and if victims and offenders were equally represented in READI, we would expect there to be between 9 (1/.11) and 20 (1/.05) shooting victims for every shooting arrest. As shown in Appendix Table A.IX, that is exactly what we see: the control group has just over 13 shooting victims for every shooting arrest. On the victimization side, outside research suggests that about 1 in 5 firearm injuries result in a homicide (Cook et al., 2017). When we disaggregate our victimizations into shootings and homicides, 23 percent of the shooting and homicide victims in the control group are homicides. In both cases, our definition of non-fatal shootings generates statistics that match what we would expect from outside data.

A.2.3 Incarceration data: jail and prison

We use two sources of incarceration data: records from the Cook County Sheriff's Office (CCSO) capturing detainees in the Cook County Jail each day from 2015 to present, and records from the Illinois Department of Corrections (IDOC) capturing entry and exit dates in state carceral facilities from 1999 to present. CCSO records are linked to study members using the same fingerprint-based IR number used in CPD's arrest records. We used probabilistic matching based on name and DOB to match IDOC records to our study sample.

We use these two data sources to generate baseline covariates for the number of days incarcerated in the 30 months prior to randomization and measures of incarceration during the 20-month outcome period. We also used these records during study recruitment to filter out individuals prior to randomization who were incarcerated at the time they were referred to READI. However, since we only received periodic updates of the IDOC data—annually during the first half of the study and bi-annually during the second half—some individuals who had only recently become incarcerated were nevertheless included in randomization.

A.2.4 Data from READI providers

We rely on multiple datasets from the READI partner organizations to measure program take-up, as well as participants' hours and earnings.

Our primary source for measuring take-up are the monthly headcount data from Heartland Alliance. The headcount data contain the first and last names of everyone who participated in READI during a given month. To distinguish between participants with the same first and last name, as well as to determine the date (rather than month) they took up, we supplement the headcount data with information from two additional sources: records from the three READI employment organizations and weekly payroll data from Heartland, the employer of record for READI participants. The employment organization data include information on a participant's full name, DOB, and date of orientation. For participants who do not appear in the employment organization data, we use the date when they are first paid in the weekly payroll data as their take-up date. Note that the payroll data only include information on participants who both signed a consent form allowing their records to be shared with researchers and who started the job; 117 men started the job prior to consent forms being distributed, and 7 men attended orientation but did not start the job within 20 months of randomization.

We measure participants' hours and earnings separately for work and for CBT and training sessions. For work hours and earnings, we use the weekly payroll data, which contain information for all but the 124 participants described above. For CBT and training hours and stipends, we use records from the READI employment organizations. Prior to COVID, these data consist of signatures on sign-in sheets, many of which are not legible. Ultimately, 13 months of sign-in sheet data between July 2018 and July 2019 from two of the three organizations contained signatures legible enough for us to hand-code participation. During this time period, participation in CBT and training sessions is very closely correlated with when people appear in the payroll data. We use these data to extrapolate CBT and training hours and stipends during the pre-COVID period. Specifically, we limit to the participants who appear in the payroll data; calculate the average number of CBT/training sessions they attend per week; multiply the average number of weekly sessions attended by the number of weeks they were paid; and then multiply the estimated total number of sessions attended by 1.5 hours (to get the total hours attended) or $25 (to get the total stipend amount). From April 2020 onward, we received detailed Excel spreadsheets from the organizations containing information on weekly CBT and training attendance. Note that, because percentiles of the CBT/training hours distribution cannot be scaled across missing data the way that averages can, we do not report these percentiles in Table III.

A.2.5 Arrest data & bias

As discussed in Section 3.1, the key assumption for treatment–control differences in arrests to successfully proxy for treatment–control differences in offending is that treatment does not change the probability of arrest conditional on actual criminal behavior. There are two ways this assumption might not hold. First, program-driven changes in time use may affect the probability of interacting with a police officer. For example, if treatment decreases idle time spent outside, it might lower the probability police happen to notice someone who has an outstanding warrant, or that they arrest someone for a crime they did not commit (e.g., for loitering or as part of a general round-up). Second, conditional on an interaction, treatment may affect the likelihood of talking one's way out of—or into—an arrest. For example, how a person responds to an officer could determine whether a street stop results in a disobeying an officer or disorderly conduct arrest.

There are several reasons to think these issues are unlikely to be a major problem in our setting. First, if improving skills or employing people during the day actually reduced the probability of arrest, we would expect to see consistent arrest declines in other programs that accomplish these changes. But many jobs programs have shown, on average, no decline in arrests (Bloom et al., 1997; MDRC, 1980; Redcross et al., 2016), and there is evidence from surveys that while education does decrease crime, it does not change the probability of arrest conditional on crime (Lochner, 2004; Lochner and Moretti, 2004). Second, if READI reduced police interactions or improved the quality of the interactions, we would expect those changes to be most salient for the kinds of arrests that involve the most police discretion. Yet as we show in Appendix A.5.7, we can rule out even relatively small declines in our "other" arrest category, which includes most discretionary arrests such as vandalism, trespassing, loitering, disorderly conduct, disobeying an officer, and so forth. Combined with the increased scrutiny that serious violence charges like shootings and homicides entail from prosecutors following an arrest, it seems quite unlikely that READI is substantially changing the probability of being arrested conditional on committing (or not committing) a serious violent crime.

A.3 Eligibility and randomization details

Eligibility for READI was limited to men 18 and over at the highest risk of gun violence involvement who lived or spent considerable time in five Chicago neighborhoods (Austin, West Garfield Park, North Lawndale, Englewood, and West Englewood). There were three main reasons for the age floor. First, in the year prior to READI, people 18 and over made up 87 percent of shooting victims. Second, since READI offered full-time work, excluding youth was a way to avoid crowding out high school attendance. And third, conversations with city violence-prevention staff and our own review suggested that most of the city's existing violence prevention programs served youth, leaving a service gap for the population comprising the vast majority of shooting victims. As described below, some of the referral pathways into READI also set an age ceiling of 40. READI focused on men because they are disproportionately involved in gun violence; because the most frequent forms of lethal violence among women have a different set of causes and would likely require a different kind of intervention; and because there were not enough resources to operate two different versions of the program. Although race was not an eligibility criteria, in practice nearly all men in the study are Black (Table I). This is largely a function of the racial homogeneity of the READI neighborhoods, though, as Heller et al. (2022) show, Black men in Chicago have a disproportionately high risk of being shot.

Randomization occurred between August 2017 and March 2020. Because of the onset of the COVID-19 pandemic, we stopped randomization, resulting in a study sample approximately 20 percent smaller than initially intended. Intake at a given time was based on provider capacity, determined by funding levels and the current caseloads of READI outreach workers. We used all available data at the time of randomization to screen out those who had either been incarcerated or died since their referral, though we received carceral and homicide data with a lag, so the screening was imperfect. Below we describe the referral and randomization process for each referral pathway, since processes differed based on the source of the referral.

The main text reports a joint test of significance of baseline covariates showing that randomization successfully balanced observable characteristics, and Appendix Table A.I reports a similar test by pathway. The covariates included in the baseline test, in addition to randomization strata fixed effects, are: age; the baseline versions of our three primary index components (counts of arrests for shooting and homicide, victimizations for shootings [no baseline homicide victims], and arrests for other serious violence); counts of arrests for less serious violence, property, drug, and other crimes; counts of non-shooting and non-violent victimizations (separately); days incarcerated in the prior 30 months; an indicator for being incarcerated at baseline; the predicted risk of future gun violence (risk score with zero imputed for missing); an indicator for missing risk score; and an indicator for not being Black (other race/ethnicity or missing).

A.3.1 Algorithm referral pathway

Referrals from the algorithm pathway, which began in December 2017 and occurred over 13 rounds through October 2019, started with the output from a machine learning algorithm trained to predict gun violence involvement in the next 18 months. Because READI recruited on a rolling basis, and because the relationship between the predictors and gun violence involvement may not be static, new models were trained 10 times over the referral period using the most up-to-date records available from CPD. To be in the prediction sample, a person had to have at least one arrest or at least two victimizations in the data during the 50 months prior to the prediction date. Excluding those with a single victimization record serves to remove a large number of people at very low risk of future gun violence (e.g., tourists who report a mugging). It also improves data quality: Non-shooting victimization records do not always have reliable dates of birth, so single cases in victim records can involve considerably more matching error. Heller et al. (2022) provides a detailed description of a related prediction model.

One key difference relative to Heller et al. (2022), which focuses solely on shooting victimization, is that the READI model trained gradient-boosted decision trees to predict a broader outcome: whether someone would be either arrested for, or the victim of, a violent crime involving a gun during the next 18 months (to match READI's basic service period).[5a] Violent crimes involving a gun included homicide, assault, battery, or robbery with a firearm.[6a] We included arrests here to proxy for gun violence offenses like homicides, shootings, and armed robberies, despite concerns about arrests capturing both individual behavior and police decision-making. The reason is that those in charge of READI were very interested in serving potential offenders as well as potential victims. Since READI provided what was perceived as a valuable opportunity rather than punishment or police involvement, the costs of using information influenced by potential police bias to refer men to READI seemed relatively low. Additionally, in practice, arrests are a very incomplete measure of such behavior due to the low arrest or "clearance" rate in Chicago for these offenses. For example, in 2016, only 26 percent of homicides and 5 percent of non-fatal shootings in Chicago resulted in an arrest (Kapustin et al., 2017). As a result, the large majority of the outcomes we predict are gun violence victimizations (91 percent).

The model used over 1,400 features built from arrest and victimization records to predict future violence. The model continued to develop over the course of READI. At the beginning of the study, we did not have permission to use either juvenile records or information on domestic violence victimizations to train the model or generate predictions. We obtained permission to use juvenile records in February 2019. Other refinements were made over time as we continued to learn from studying the prediction model. We never used race or ethnicity as predictors in the model, since there are open legal questions about the role of such features in determining service provision.

The output of a given model is a set of predicted probabilities of gun violence involvement in the 18 months following the model's prediction date for around 300,000 people. To identify potential candidates for randomization and referral to READI, we imposed four restrictions on the output of the model. First, we limited our focus to adult men under the age of 40. Second, we limited our focus to men with a high likelihood of living in one of the neighborhoods where READI operates (identified from the home address and incident locations in the CPD data). Third, because the referral process involves sharing information about a person derived from confidential police records with an outreach organization, we excluded anyone about whom publicly available information was unavailable. In practice,

[5a] Included in the outcome is whether an individual appears in CPD's homicide offenders dataset. As discussed in Appendix A.2.1, there is imperfect overlap between the arrests and homicide offenders datasets.

[6a] The predicted outcome excludes suicides and incidents where someone is shot or killed by a police officer.

this meant limiting to only men with at least one adult (18+) arrest since January 1, 2014, as these arrest records are public.[7a] Finally, we removed anyone who died in the period after their predicted risk was generated, or who was incarcerated in an IDOC prison or in the Cook County Jail when randomization and referral were about to occur. After imposing these restrictions, men not already in the study were ranked in descending order of their predicted probability of future gun violence involvement, separately within each of the three READI sites.

If a program provider needed X individuals for a given randomization round, we selected the 2X individuals with the highest predicted risk in their site. We then randomized this group, so everyone had a 50 percent chance of being assigned to a READI offer.[8a] Once the treatment group was selected, we compiled publicly available information on each individual (name, age, mugshot, and locations of recent arrests) into a referral sheet and provided these to the outreach organizations.[9a]

Upon receiving these referral sheets, workers at the assigned READI outreach organization used their social networks, social media, and public databases to locate the individuals. This process could take weeks or months. Even after the search process, a significant share of referred men could not be found: according to self-reported data by outreach workers, of the 775 men referred through December 2018, 14 percent could not be located. For those men who could be found, outreach workers would explain the nature of READI, attempt to develop a relationship, and persuade them to join. Those who refused were still eligible to enroll at a future date, and outreach workers often remained in close contact.

The primary way in which the algorithm served READI was by acting as a standalone referral source. However, the algorithm may have also affected the composition of referrals from the two other pathways. While the algorithm was not used to screen in or out any outreach referrals, the outreach organizations were aware that Heartland Alliance monitored the average level of predicted risk of the men whom they referred.[10a] Anecdotally, we heard

[7a] `http://publicsearch1.chicagopolice.org`

[8a] See Appendix A.3.4 for exceptions to this.

[9a] In the beginning of the study, the referral sheets included only a referral's name and age. This limited information proved very difficult for outreach workers to use successfully, since many of the men being referred did not go by their legal names, nor was there any indication of where an outreach worker might find them. We then added mugshots and locations of recent arrests to the sheets, both of which were publicly available via CPD's online adult arrest record search. Starting in February 2019, to further assist outreach workers in finding referrals, we had research staff go to the courthouse in Chicago to use terminals that contain additional information such as referrals' three most recent home addresses and any upcoming court dates, which were added to the sheets. Outreach workers reported that this additional information helped them locate and recruit these individuals.

[10a] The risk scores of individual outreach referrals were not shared with Heartland Alliance or even the outreach organizations who referred them, to protect their privacy.

that this use of the algorithm as an accountability tool may have altered whom outreach organizations referred to READI for randomization. In addition, among a subset of re-entry referrals—those from Division 6 of the Cook County Jail—the algorithm was used to screen out anyone below the 95$^{\text{th}}$ percentile of the predicted risk distribution (see Appendix A.3.3).

It is also important to emphasize that, while the algorithm was trained using data from the CPD, neither the algorithm itself, the individual risk scores it generated, nor any information about individual study members were ever shared with CPD, any other law enforcement organization, or any other third party. The sole purpose of the algorithm is to provide local organizations such as those implementing READI with publicly available information about people who may be at high risk of involvement in gun violence in the communities they serve.

A.3.2 Outreach referral pathway

The second referral pathway relies on the experience and knowledge of front-line staff at READI's partner outreach organizations to identify men currently involved in gun violence in their neighborhood. These outreach organizations are community-based non-profits with longstanding roots in their neighborhood, and they typically have been involved in a range of violence-reduction activities and youth programming for some years. Their front-line workers are recruited from the communities they serve and often have backgrounds similar to those of program participants, including exposure to secondary trauma (Hureau et al., 2022). This experience provides them with extensive information networks, contacts, and the familiarity and credibility to approach high-risk young men. The partner organizations employed READI-dedicated outreach workers to search for, identify, screen, and recruit READI participants.

Outreach workers were instructed to use community contacts and their organization's sources to identify the men 18 and older whom they believed to be at highest risk for being involved in gun violence in the coming months. No age cap applied to outreach referrals to give maximum discretion to human expertise, though in practice, only 5 percent were over 40 at baseline. After identifying someone as a potential candidate, an outreach worker discussed the possibility of READI participation with them, explained the study, and administered a 10-question risk assessment. The questionnaire asked whether the person was a victim of a violent crime; had previously been incarcerated for a gun-related offense; was an active member of a street gang; had substance abuse issues; was unstably housed; was promoting violence on social media; and had recently been arrested. Outreach workers were instructed that a referral had to meet, at a minimum, more than one of these conditions. Outreach organizations then discussed the set of interested candidates and decided whether to nominate them for random assignment.

Each time a program provider was ready for new outreach referrals, they provided us with an even-numbered list of people who had been through this screening process. The research team then vetted the list, returning anyone whom we either could not locate in available administrative data or had been previously randomized.[11a] The provider then either corrected identifying information for those whom we could not find or provided new names to replace anyone previously randomized. If the resulting number of candidate outreach referrals was greater than twice the number of open program slots, we randomly dropped candidate referrals to obtain a group exactly twice the size of the number of slots. Unused candidate referrals were returned to the outreach organization so that they could be referred in a future round.

Once we had a vetted list with exactly twice the number of names as available slots, we randomized with a 50 percent treatment probability. We then returned the list of selected individuals to the providers so they could start the intake process. This referral process occurred 57 times, from August 2017 through March 2020.

The outreach referral process creates several key differences between outreach referrals and algorithm referrals. First, because outreach workers typically knew the location and interest of a person before randomization, we expected (and later confirmed) higher rates of take-up for men referred through this pathway. Second, because outreach workers knew whether a person they referred was randomized *not* to receive a READI offer, they may have worked to find such a person alternative programming. In contrast, a person identified by the algorithm pathway and randomized to the control group was not contacted by the research team, nor was their identity shared with outreach, making it less likely that they were referred to alternative programming. Lastly, while the algorithm pathway mechanically identified men solely on the basis of their risk of future involvement in gun violence ($\hat{Y}(0)$) as predicted using CPD records, outreach workers could have identified men based on some combination of risk of future gun violence involvement ($\hat{Y}(0)$) and being ready and willing to participate in programming (β), among other potential factors.

A.3.3 Re-entry referral pathway

The third and final pathway focuses on men being released from incarceration, including those released on condition of parole from an IDOC prison facility, and those exiting the Cook County Jail (including via the use of electronic monitoring). Similar to outreach referrals, this pathway relies on the knowledge and experience of implementation partners, in this case parole officers and others working in detention facilities. They were told about

[11a]We attempted to match each name and date of birth to our administrative records using both exact and probabilistic matching. If someone was not found in CPD records, we also referenced Chicago Public Schools data to try to identify the individual.

the population READI aimed to serve and asked to refer those at high risk of gun violence. As with outreach referrals, location and interest in program participation were typically ascertained prior to nomination for random assignment for this pathway.

The re-entry pathway consisted of referrals from three different carceral settings. First, for people leaving IDOC custody to parole, parole agents received a set of criteria developed to aid in identifying READI's target population of high-risk men, including having been a victim or offender in a violent or gun crime, as well as being over 18 and living or being paroled to a READI neighborhood. The agents shared those referrals in batches with the research team, at which point the process looked identical to the vetting and randomization for outreach referrals.

Second, at the Cook County Jail, an embedded READI staff member identified eligible individuals from among those housed in Division 6 or participating in the Sheriff's Anti-Violence Effort (SAVE), which serves men 18-24 years old from Chicago's 15 highest-violence neighborhoods.[12a] To be eligible, men had to be between ages 18 and 40, live in READI zip codes, and have a predicted probability of future gun violence involvement using the algorithm described above that was at or above the 95$^{\text{th}}$ percentile. Once this group was identified, Heartland Alliance staff consented eligible men while they were in the jail. Because release dates were unpredictable, consented men did not become part of the study and were not randomized until they were released. The first set of referrals occurred in a group, but after that men were randomized one-by-one at the point of release, with a treatment probability of 50 percent. Those assigned to the treatment group were then assigned to outreach workers based on the neighborhood to which they were returning.

Lastly, a small group of referred men entered the custody of the Illinois Department of Juvenile Justice (IDJJ) as youth but were exiting custody as adults. Individuals leaving IDJJ custody enter Aftercare, the community supervision program run by IDJJ that is analogous to adult parole. Aftercare specialists referred people 18 or older living in READI neighborhoods to Heartland Alliance staff, who then met with those individuals to recruit and consent them. If the referred individuals wished to participate, READI shared their information with the research team for random assignment, which occurred either in pairs or one-by-one, with a treatment probability of 50 percent.

[12a]Division 6 of the Cook County Jail houses minimum- and medium-security individuals. Following conversations with jail administrators about the population READI was designed to serve, the need to identify individuals likely to be released from the jail (rather than go on to serve a prison sentence), and logistical constraints to in-facility recruitment, it was determined that Division 6 was the most promising source of eligible individuals.

A.3.4 Variation in treatment probability within strata

There are nine cases where people who were ineligible for READI based on their baseline characteristics slipped through our initial screening process but could later be identified: two women, three people who were deceased by the date of randomization, and four duplicates who had already been randomized. Since these characteristics are observable for both treatment and control groups, we can drop these nine cases from the study without undermining randomization. Some of the resulting randomization strata have an uneven number of people in them, with treatment probabilities not exactly equal to 0.5. To account for this aspect of the experimental design, we control for randomization strata fixed effects in all analyses. Re-entry referrals recruited and consented in the jail were only randomized upon discharge, which often occurred one-at-a-time rather than in groups. We group everyone in this pathway who was randomized individually into separate strata based on their quarter of random assignment; each person had exactly a 0.5 probability of being assigned to a READI offer, but the realization of multiple Bernoulli trials means these strata also have some variation around 0.5.

A.4 Program details

A.4.1 Jobs and wages

READI jobs were designed to help participants with very little work readiness and multiple barriers to employment stay involved in the program. Men who stopped attending work for a time were allowed to later resume participating. If a participant disengaged from work, or committed an offense on site, outreach and program staff continued to engage them. Participants had opportunities to return to the program after the problems had been resolved, with restorative justice processes used to reintegrate any violent offenders back into the program.

READI's jobs and wages were designed in stages with progressively broader goals set for each stage. Appendix Figure A.I shows the initial conception of how these stages would work, along with how wages would grow as participants advanced across stages. From the time of its conception, READI's designers knew that relatively few participants could be expected to make it all the way to Stage 4. This was not a typical "transitional jobs" program, where the goal was to establish long-term employment. Indeed, increasing employment was not a primary outcome of the intervention; the focus was always on preventing violence.

Appendix Figure A.II reflects how wages actually grew over time, with the number of weeks after initial take-up on the x-axis and change in wages on the y-axis (the graph adjusts for the rising minimum wage in Chicago, so it shows only progression through stages with increased wages). The average participant who persisted in the program earned 40–50 cents more per hour by the end of READI than at the beginning, which is about 1 additional stage.

A.4.2 CBT

In READI, participants took part in 90-minute group CBT sessions three mornings per week. Facilitators delivered a version of the University of Cincinnati's Cognitive Behavioral Interventions Core Curriculum (CBI-CC) modified to be culturally relevant and targeted to participants' literacy levels. Staff also incorporated aspects of the Seeking Safety curriculum to address substance abuse and symptoms of traumatic stress. Sessions were co-facilitated by a senior coach and an experienced CBT facilitator who is a member of the READI program staff. Cross-community meetings of CBT staff aimed to ensure that the model was being delivered consistently across READI's sites, though in practice there was considerable variation across providers and staff members, as well as over time.

Participants received $25 gift cards for each CBT and personal development session they attended. These gift cards were for specific vendors, including stores such as Walmart and Kroger, fast food chains such as McDonald's and Subway, the Chicago Transportation Authority, and retail stores such as Foot Locker.

A.4.3 Safety during READI programming

Maintaining participant and staff safety during programming was a major focus of program operations. To avoid reports about participant locations to non-participating rivals, READI participants were banned from having their cell phones at their worksites. They were also forbidden from carrying weapons, enforced by metal-detecting wands. READI provided group transportation from morning CBT sessions to the worksites, to avoid the dangers involved in crossing gang lines. Staff would also sometimes rearrange work crews to help manage ongoing personal conflict and group rivalries. READI program operators learned over time what was more and less successful in ensuring safety, and program rules continued to develop over the course of the program. For example, the crew chiefs on worksites would sometimes call the outreach workers to help resolve conflicts while participants were at work. And worksite assignments were rotated so that no one could consistently anticipate where participants would be.

A.5 Additional results

This section provides further discussion and details of results referenced in the main text.

A.5.1 Covariates, functional form, and inference

The main text uses the same set of baseline covariates for all regressions. To minimize any finite sample bias stemming from mis-specification, we specify baseline covariates as a set of indicator variables (dividing continuous variables into quartiles, with 0 as a separate category, and dividing count variables into an indicator for 0 (left out), an indicator for 75$^{\text{th}}$ percentile and above, and then dividing the remaining variation into roughly equal groups). The indicators are for: at least one arrest for a shooting or homicide; 1 or more than 1 shooting victimization; 1 or more than 1 other (non-violent) victimization; 1, 2-3, or more than 3 non-shooting (violent) victimizations; 1 or more than 1 other serious violent-crime arrests; 1-2 or more than 2 less serious violent-crime arrests; 1-2 or more than 2 property crime arrests; 1-2, 3-4, 5-8, or more than 8 drug crime arrests; 0-3, 4-7, 8-11, or more than 11 other arrests; quartiles 1, 2, 3, and 4 of days in jail during the past 30 months (0 as baseline category); quartiles 1, 2, 3, and 4 of days in prison during the past 30 months (0 as baseline category); quartiles 2, 3, and 4 of age; quartiles 2, 3, and 4 of baseline risk score, and an indicator for missing baseline risk score; and an indicator for having a race/ethnicity other than Black (including missing).

To ensure that results are not sensitive to the way covariates are included or functional form, Appendix Table A.IV reports several sets of alternative results. The first panel reports the estimated effect for a version of the primary index that pools arrests for shootings and homicides and other serious violent-crime arrests into a single component, as described in our pre-analysis plan. The result is very similar to the one obtained using our preferred version of the primary index that separates these two arrest types. The second panel shows estimated effects for the main outcomes without any covariates, other than the randomization strata fixed effects. Although the results are slightly less precise, as expected when excluding covariates that explain some of the residual variation in the outcomes, they are nevertheless substantively quite similar to those referenced in the main text. The third panel shows results with covariates selected by the post-double selection LASSO (Belloni et al., 2014a,b); note the randomization strata fixed effects are always included. The LASSO selects the indicator for having a missing risk score in the index regression, and does not select any covariates for any of the three index component regressions.

The fourth panel reverts to a specification based on what we understood to be available at the time of our pre-analysis plan. At that time, we did not expect there to be variation in treatment probabilities across randomization rounds. As a result, we pre-specified site and pathway fixed effects rather than strata fixed effects (which combine site and pathway with the randomization date), but mentioned we would use inverse probability weighting (IPW) if there were any variation in treatment probabilities. So the third panel implements this version with IPW. In addition, our pre-analysis plan inadvertently left off information about baseline risk score and race/ethnicity, which we also exclude from the covariates in this panel. We enter the other covariates—age, shooting victimizations, non-shooting (violent) victimizations, other (non-violent) victimizations, Part I arrests, less serious violent-crime arrests, property crime arrests, drug arrests, other arrests, and days incarcerated (which, as pre-specified, we did receive and establish to be useful as a measure while blinded to treatment)—in levels rather than as indicators.

Finally, since the main components measures are count variables, the last panel shows results using a Poisson specification, with robust standard errors to relax the assumption that the mean equals the variance. To be comparable to the ITTs reported in other panels, we show the average marginal effects from Poisson models that include the standard covariates (excluding strata fixed effects to ensure models converge). Across all the panels, the results are quite similar.

A different check on our results is to ensure that an alternative kind of inference—Fisherian randomization inference—generates the same conclusions. Randomization inference tests the sharp null that there are no treatment effects for anyone (unlike the heteroskedasticity-robust standard errors in the main text, which test the null that the average effect of treatment is zero). Randomization inference has the additional benefit of estimating the distribution of the key test statistic, the regression-adjusted difference of means between treatment and control groups, under the null without additional assumptions about the distribution of the errors or the independence of outcomes across observations. Appendix Table A.V compares the p-values from the main text to those estimated via 5,000 iterations in randomization inference. In all cases, the conclusions are exactly the same as inference about the average effects.

A.5.2 Scaling for incapacitation

Our main outcome measures are counts of serious violent-crime arrests and victimizations within the 20-month outcome period. The fact that people may be incapacitated and therefore unable to be involved in a serious violent crime, however, means that these counts capture a combination of individual behavior and "incapacitation," in this case either through incarceration or death. We pre-specified our interest in counts as outcomes, because in some respects the number of incidents are what matters most for the social cost of violence; if someone in the control group is incarcerated for the full outcome period, they are unable to engage in violence. As long as we account for the social cost of incarceration in our benefit-cost comparison, we can ask whether the benefits of READI outweigh its costs.

As we noted in our pre-analysis plan, however, there is an additional question of substantive interest: whether READI changes behavior during the periods when someone is free to choose how to spend their time. In some respects, the results in Appendix Table A.X, discussed below, suggest that adjusting for incapacitation should not matter much, since there is no significant difference in the number or percent of days the treatment and control groups are incapacitated. But it is possible that treatment heterogeneity could be masking differential incapacitation changes that affect the serious violence results.

Testing for program effects conditional on incapacitation requires a different analysis strategy. Survival analysis is a common way to assess whether time to initial incidents changes. But in our case, survival analysis is complicated by competing risks: we are most interested in serious violence outcomes, but individuals' data can be censored by other types of failures (i.e., incarceration for more minor arrests or probation/parole violations). Since these types of censoring are not ignorable—it is plausible READI could affect them—standard survival analysis methods cannot help us to isolate changes in serious violence involvement alone.

An alternative is to scale the observed counts of violent incidents to account for the number of days an individual is neither incarcerated nor deceased ("is available"). We do this by using rates as outcomes (number of incidents/number of days available), where the number of days available is the number of days in our 20-month outcome period (610) minus days incarcerated minus days after a homicide victimization. It is likely that some individuals with non-fatal shooting injuries or other serious health issues were also incapacitated in the hospital for some amount of the outcome period. But we do not observe hospital stays, and there are incidents of serious violence in our data that occur within days following a non-fatal shooting victimization. So we do not adjust for any unobserved incapacitation due to factors other than incarceration (in prison or in jail) or being killed.

A key challenge when using rates as a dependent variable is the leverage created by a small number of outliers: those who were incarcerated long-term or killed almost immediately after randomization (i.e., that have a very small denominator), and therefore had few days available during the outcome period. To reduce the influence of these outliers, we report two winsorizations of the rates by top-coding using the 99.5^{th} and the 99^{th} percentile of the rate distribution. Note that our rate variable is at the daily level. To facilitate comparisons between the rate results and our main count results, we multiply all coefficients and control means by the number of days in the data (610).

The results are in Appendix Table A.VI. Rates are scaled up to the full 610-day outcome period (i.e., multiplied by 610) to be comparable in magnitude to the count outcome over 20 months. It is clear that concern about the role of outlier rates is merited. Across the three outcomes, the control means when using raw rates as the dependent variable are between 1.7

and 3.2 times as large as the control means for counts, and standard errors grow by factors ranging from 2.3 to 11. As a result, we do not put much stock in the results for raw rates.

Focusing on the two winsorized rates, there is some evidence that the insignificant decline in the count of shooting and homicide victimizations masks a significant decline in the rate of these victimizations. Both winsorized rate estimates are much larger than our main estimates, suggesting declines of between 8 and 9 for every 100 compliers (41–45 percent reductions, unadjusted $p = 0.03$ and $p = 0.05$). This pattern is what one might expect if treatment lowered the rate of being shot, and as a result kept participants alive and available to be at risk of victimization for more time than controls. Appendix Table A.X below does show a marginally significant decline in days lost to homicide. While this result may suggest that risky behavior is falling more generally, not just offending behavior, the result would still not survive adjustments for multiple testing. So as in our main analysis, future replication would be very valuable.

Results for shooting and homicide arrests are less sensitive to using rates rather than counts. Across all versions, point estimates vary very little. IV estimates range from -0.0226 to -0.0297, compared to the count version of -0.0220 (57–59 percent declines, compared to 65 percent for counts). Statistical significance is more variable, as the skew in the rates increases standard errors. But winsorizing at the 99$^{\text{th}}$ percentile is still marginally statistically significant (unadjusted $p = 0.09$).

Rate results for other serious violent-crime arrests are more variable, with point estimates and standard errors varying quite a bit across the different versions of rates. But none of the results approach statistical significance, so the conclusion of a positive point estimate indistinguishable from zero does not change when accounting for differential days available.

A.5.3 COVID-19

As the main text describes, the onset of the COVID-19 pandemic brought dramatic changes to program delivery. Supported work stopped entirely from mid-March 2020 until August 2020, since in-person operations were not safe to continue. To prevent simply cutting off current READI participants, program operators continued to provide participants with "standby pay" equal to the average amount that they had earned over their last 3 weeks prior to the onset of the pandemic. CBT sessions shifted online, which likely diminished their quality given internet connection struggles, challenges finding privacy to discuss personal issues, and safety issues involved with potential disclosure of a person's location to members of rival groups (without the safety safeguards that READI operators had in place). The period after the start of the pandemic also saw changes in the crime environment in Chicago, with many types of crime and arrests decreasing as people stayed home, but a huge spike in gun violence. Because recruitment began in August 2017 and our outcome window is 20 months, approximately 76 percent of post-randomization person-day observations occurred before the onset of the pandemic.

Appendix Figure A.III shows program retention as in the main text, but here including data from the COVID-19 period. The patterns are generally similar, with somewhat more fall-off in program participation during the end of the program, as COVID-19 deterred some individuals from returning. Appendix Table A.III shows participation details separately for pre- and post-COVID-19 periods.

Appendix Table A.VII reports READI's estimated impact separately in the pre- and post-COVID periods. We are largely under-powered to test the difference between the two periods, although the decline in shooting and homicide victimization is suggestively different (unadjusted for multiple testing), with a much larger protective effect post-COVID. This change seems to be unique to potentially lethal victimizations, though, as the point estimate on all victimizations during COVID-19 is actually positive.

A.5.4 Spillovers

A priori, it is not clear what impact spillovers would have on the interpretation of our point estimates. If exposure to treated peers has the same effect regardless of one's own treatment status, then exposure should be balanced across treatment and control groups. In this case, the ITT might miss the net social effects of READI by failing to account for level shifts in both treatment and control groups. But the ITT would still be an unbiased estimate of the average difference between being offered treatment or not, inclusive of social spillovers. If being exposed to treated peers interacts with one's own treatment status, however, then the ITT could either over- or understate the net effects of READI. For example, if members of the control group pick up the guns that the treatment group puts down, then the displacement would lead the ITT to overstate the net benefits of READI. Alternatively, if the control group learns some of the treatment group's positive behaviors, or if the drop in violence among treated men cools off what would have been an escalating cycle of violence in a neighborhood, then the ITT may understate the net benefits of READI.

As discussed in the main text, the evidence in Craig et al. (2022) provides some weak evidence that the ITT effect on serious violent-crime arrests that are not shootings or homicides may be an inflated measure of READI's direct effect (i.e., that the actual direct effect on these kinds of arrests could be negative despite the positive ITT estimate, due to differential spillovers onto treatment versus control peers of treated individuals). But the clearest evidence of a SUTVA violation is for drug-crime arrests, where it appears that READI actually decreases these arrests both among those directly treated and among controls who are exposed to treated peers.

A.5.5 Dynamic treatment effects

Appendix Figure A.V shows how treatment effects accrue over time during the 20-month outcome period. The top left panel shows cumulative participation rates, and the other panels show cumulative ITT effect estimates. The estimated decline in shooting and homicide arrests accrues over time, becoming statistically significant at about the year mark. The estimated decline in shooting and homicide victimization, while noisier, follows a similar pattern.

A.5.6 Incapacitation as mechanism

To test whether keeping people busy during the workday mechanically reduced violence during that time, we separately code violent incidents based on the day and time they occur (measuring the time of the incident rather than the time of an arrest). Appendix Table A.VIII shows results separately for incidents that occur during the work day (8am–6pm Monday through Friday), weekend (Friday 6pm–Sunday 11:59pm), and weekday mornings and nights (Monday through Friday 6pm–8am). While point estimates on the total number of arrests and victimizations are negative and substantively large during work days, there is no indication that declines in serious violence involvement are concentrated during the work day.

Instead, the decline in arrests for shootings and homicides are driven by declines in weekend incidents, which is when most of these incidents occur. Though the small number of violent events in each cell makes this analysis noisy, all point estimates for weekend violence are negative, and the arrest decline is statistically significant ($p = 0.02$). We conclude that it is not the incapacitation effect of READI activities that decreases shooting and homicide arrests. Rather, the treatment group appears to change its behavior outside the work day. Appendix A.5.7 discusses incapacitation for outcomes other than our main violence measures.

A.5.7 Other outcomes

The three main components of our primary index are shooting and homicide victimizations, shooting and homicide arrests, and other serious violent-crime arrests. Appendix Table A.IX separates each of these components into smaller categories, with the top panel showing counts and the bottom panel showing indicators for whether the outcome ever happened during the outcome period. We did not design the study to have enough power to separately identify effects on these smaller subcategories; homicide arrests are the only category with statistically significant effects, ignoring multiple testing adjustments. Nonetheless, the table is useful for showing what is underlying the effects reported in the main text.

Homicides and non-fatal shooting victimizations, homicide arrests, shooting arrests, and (non-shooting) aggravated assault and battery arrests all have negative point estimates. The reason other serious violent-crime arrests has a positive point estimate is that the robbery and sexual assault arrests point estimates are positive and proportionally large. The control means reflect lower clearance rates for these less serious crimes, with more homicide arrests than robbery arrests (0.023 versus 0.020), despite robberies being much more common. The result has too little power for us to be confident in whether these arrests are actually increasing. Results are quite similar in the bottom panel, which uses indicator variables rather than counts as the dependent variables.

These measures of serious violence involvement were our primary outcomes of interest. But we also pre-specified a secondary interest in other types of crime, as well as incarceration (and, for understanding potential censoring, days lost to homicide). Appendix Table A.X shows these results. A few results have large enough point estimates to approach statistical significance on their own, with p-values of 0.1 and 0.17 for a decline in any drug-crime arrests and in days incapacitated, respectively, and with a more precise decline in days lost to homicide (consistent with the large negative point estimate in shooting and homicide victimizations). Given the number of tests in this table, however, the general take-away should likely be that READI did not seem to affect these other measures of criminal involvement.

Appendix Table A.XI breaks down the non-primary arrest and victimization outcomes by time of day. As mentioned in the main text, there is some indication that work day incapacitation plays a role in reducing drug-crime arrests; during the work day, for every 100 people there are 4.6 fewer arrests for drug crimes among the treatment group than the control group (a 27 percent decline, unadjusted $p = 0.02$). It is possible that some of the work day drop in drug-crime arrests has to do with policing strategies. If police make arrest sweeps of drug markets during the day on weekdays, READI participants may be less likely to be caught up in those sweeps even if their involvement in drug crime moves to other times. However, point estimates on drug-crime arrests at other times are either very small (weekend) or negative (weekday morning and night). So the work day incapacitation does not seem to be purely a shift in when drug crimes and police enforcement happen. Other (non-shooting) violent victimizations also fall during the work day by 3 per 100 people (28 percent, unadjusted $p = 0.05$), consistent with the possibility that READI keeps the treatment group safer during the work day.

The only indication of significant substitution of criminal behavior to other times is an increase in less serious violent-crime arrests during the weekend (2.1 additional arrests per 100 people assigned to treatment, a 34 percent increase, unadjusted $p = 0.08$). Although this is balanced by negative point estimates at other times of the week, it could be consistent with the basic pattern of substitution from more to less serious violence seen in the primary index components.

A.6 Heterogeneity

A.6.1 Other pre-specified subgroups

The main text focuses on heterogeneity by pathway and baseline risk, which were built into our randomization and experimental design, as well as the interaction of the two. Our pre-analysis plan also pre-specified an interest in heterogeneity by geography and age, along with an acknowledgment that all of the heterogeneity analyses would be exploratory since we would likely lack power to distinguish effects across groups.

Appendix Tables A.XII and A.XIII report results for our primary index and its components by these groups. An F-test for whether the primary index varies by group finds significant heterogeneity by site ($p = 0.09$), but not for age ($p = 0.53$). Appendix Table A.XII shows that the site heterogeneity is driven by larger declines in two sites (Austin/West Garfield and Greater Englewood), but a positive point estimate in the third (North Lawndale). We are cautious about telling too strong a narrative about these results, since many factors—program providers, population, levels of violence, and policing, among others—all varied across sites.

One potential hypothesis for future exploration, based on our qualitative work, is that when North Lawndale's program started, many of the initial treatment participants were from one particular "clique" or "crew" in that neighborhood. As recruitment broadened, ongoing conflicts in the neighborhood may have spilled over into program time. It is possible that the potentially adverse effects in North Lawndale reflect the importance of managing local conflicts as part of the program's delivery, and the difficulty of running a program with members of groups that are currently in conflict with each other.

A.6.2 Predicting outreach referrals

Given that program effects are most pronounced for outreach referrals, one way to understand that heterogeneity is to explore what observable characteristics are associated with being in the outreach pathway, and then whether those observables explain the differences in treatment effects. This section and the following take each piece of this question in turn.

Appendix Table A.XIV reports the results of regressing an indicator for being in the outreach pathway on the same set of standardized covariates we used above to predict take-up. Overall, observables explain somewhat more of the variation in referral pathway than they do for take-up decisions; the adjusted R^2 on the outreach prediction regression is 0.168. Site and most counts of arrests and victimizations are not particularly predictive of being an outreach referral, all else equal, with most point estimates being slightly negative. Outreach referrals do have significantly fewer baseline serious violent-crime arrests (that are not shootings or homicides) and days incarcerated.

The observable with the largest partial correlation is the risk score, with a one standard deviation increase in risk score being associated with a 15 percentage point decline in the probability of being an outreach referral. This is consistent with the baseline descriptive statistics in the main text, which show that outreach referrals have significantly lower predicted risk of violent gun crime than the algorithm and referral pathways. Outreach referrals are also 11 percentage points less likely to have a missing risk score. This is largely mechanical, since algorithm referrals could not have a missing risk score. Outreach referrals were also much less likely to be non-Black, which is largely a function of the fact that the partner organizations worked in disproportionately Black neighborhoods.

A.6.3 Assessing role of observables in treatment heterogeneity

Given that outreach referrals are at least somewhat observably different than referrals from the other pathways, we can ask whether those differences in observables are driving the differences in treatment effects described in Section 5.2. To do so, we follow the simple correction on observables procedure in Andrews and Oster (2021). The idea is to assess whether making the referrals from the other pathways look more like outreach referrals on observables also makes the estimated ITT effects look more similar across pathways, i.e., if observable differences explain the differences in treatment effects.

We start by estimating the probability of being an outreach referral based on observables. To do so, we first estimate a logit model of an indicator for being an outreach referral on our standard set of baseline covariates (without strata fixed effects). Next, we use the regression estimates to calculate predicted values of being an outreach referral for each individual, $\hat{p}_i^{outreach}$. We then plot the distributions of $\hat{p}_i^{outreach}$ for referrals from each pathway and use visual inspection to identify the ranges of values where there is common support—i.e., where there is sufficient overlapping density with outreach referrals in $\hat{p}_i^{outreach}$ among algorithm and re-entry referrals, respectively. When correcting for observables between outreach and algorithm referrals, this process results in restricting to referrals with $\hat{p}_i^{outreach} < 0.55$ (dropping 46 algorithm and 416 outreach referrals). When correcting for observables between outreach and re-entry referrals, this process results in restricting to referrals with $\hat{p}_i^{outreach} < 0.8$ (dropping 8 re-entry and 92 outreach referrals).

Then, we calculate for each individual $T_{is} = \frac{D_{is}}{d_s} Y_{is} - \frac{1-D_{is}}{1-d_s} Y_{is}$, where D_{is} is an indicator for whether person i in randomization stratum s is assigned to treatment, and $E_i[D_{is}] = d_s$ is the average probability of treatment assignment in stratum s. The average value of T_{is}

within a referral pathway, $\bar{T}$, is the estimated ITT. The goal is to estimate the relationship between T_{is} and the observables among algorithm and re-entry referrals, respectively, and then apply that relationship to the mean values of the observables of outreach referrals to obtain algorithm- and re-entry-specific ITT estimates that adjust for observable differences with outreach referrals. To do this, we first use OLS to regress T_{is} on our baseline covariates (without strata fixed effects) separately for algorithm and re-entry referrals. Then, we multiply the resulting coefficient estimates by the average values of the baseline covariates among the outreach referrals. Finally, we calculate the resulting adjusted versions of $\bar{T}$ for algorithm and re-entry referrals.

These $\bar{T}$ for outreach, algorithm, and re-entry referrals are reported in Appendix Table A.XV. The first column in the table reports these ITT estimates for outreach referrals prior to enforcing common support with either algorithm or re-entry referrals. The next set of columns enforces this common support condition between outreach and algorithm referrals, and therefore drops 416 outreach referrals (and 46 algorithm referrals). The negative ITT estimates for outreach referrals become much larger, consistent with the results in Section 5.3 showing that outreach referrals with higher predicted risk—those most likely to overlap with algorithm referrals and meet the common support condition—experience larger treatment effects. In contrast, the ITT estimates for algorithm referrals meeting the common support conditions prior to applying the correction procedure are mostly positive. After applying the correction procedure ("Implied ITT with Similar Xs"), the ITT estimates for algorithm referrals move even further away from those for outreach referrals. The last set of columns repeat this exercise for re-entry referrals and find the same pattern: the ITT estimates that correct for observable differences outreach referrals only widen the gap further. Both sets of results suggest that observables do not explain the treatment heterogeneity by pathway, which seems to be driven by the unobservables on which outreach workers are selecting.

A.6.4 Predicting take-up

To better understand what is correlated with participation decisions and how pathway matters for heterogeneity, we regress both extensive and intensive measures of take-up on observable characteristics, both with and without pathway indicators. For this analysis, we standardize all count covariates, so that coefficients are directly comparable in magnitude. We leave indicator covariates as indicators, with the Austin/West Garfield Park site and the algorithm pathway as the left-out categories.

The first two columns of Appendix Table A.XVI show the results for the extensive participation decision, using an indicator for ever starting the program as the dependent variable. Ignoring pathway (column 1), the strongest correlates of failing to participate (negative coefficients) are whether someone was incarcerated at the time of randomization and the algorithmic risk score. Higher-risk individuals are less likely to take up the program. There are also smaller negative correlations for those with more baseline shooting victimizations, less serious (non-Part 1) violent-crime arrests, and property crime arrests, along with lower participation in the North Lawndale site.

Once we control for pathway in column 2, however, risk score no longer predicts take-up; the difference in predicted risk is fully explained by the difference in referral pathway, though few of the other coefficients change much with the addition of pathway. In fact, pathway has the largest partial correlation with take-up by far. Both the substantial coefficients on pathway indicators and the large increase in adjusted R^2, from 0.046 without pathway controls to 0.153 with pathway controls, suggest that the unmeasured differences across pathways have a substantively important influence on the decision to take up the program.

Patterns are somewhat different when we predict the number of hours worked conditional on working. Here, the largest partial correlation shows that the small number of participants who were not Black (either belonging to another race or ethnicity or having missing race) worked about half as many hours as the average participant. The others who worked less were those with higher baseline risk scores, more property crime arrests, more non-violent victimizations, and more days incarcerated in the prior 30 months. Interestingly, although the magnitude of the coefficients on pathway indicators are in the same ballpark as these other covariates, they are not statistically significant and do little to change other coefficients when they are included. This suggests that pathway matters somewhat less for decisions on the intensive margin of work.

This descriptive analysis can help us think about what is different about those selected by humans (outreach workers or re-entry staff) relative to those selected by the algorithm. The biggest difference seems to be a much higher propensity to start the program among outreach referrals, and to a lesser extent, re-entry referrals. This pattern is consistent with the qualitative evidence about outreach workers selecting on "readiness," insofar as being ready means being willing to start participating. But as we discuss in the main text, it could also be because outreach workers have considerably more difficulty simply locating algorithm referrals to make the initial offer.

Once participants start the program, though, their propensity to keep participating is less starkly variable across pathways, all else equal. This pattern is weakly suggestive that neither the relationship between outreach workers and participants nor the potential for differential provider attention across pathways determined participant satisfaction. While hours worked is only one indication of how participants experienced READI, this take-up analysis provides one piece of evidence that outreach workers were selecting people interested in starting

READI, but once there, they were similarly invested as algorithm-referred participants.

A.6.5 Heterogeneity by predicted dosage

To aid in thinking about replicability and scale, it would be useful to know if all 18 months of the program is really necessary, or if a smaller "dose" of READI could have similar effects. The previous section showed that participants who worked more hours are observably different than those who worked fewer hours. As a result, estimating the effect of a greater dose by simply comparing outcomes across participants who participated for more or fewer hours would likely involve a fair amount of selection.

To test for whether treatment effects differ among those with high versus low levels of participation with less room for selection, we adapt the endogenous stratification exercise from Abadie et al. (2018). In that paper, the authors want to predict treatment heterogeneity over groups defined by the potential outcome in the absence of treatment, $Y(0)$. Since $Y(0)$ is unobserved for the treatment group, they develop a method to estimate the relationship between $Y(0)$ and baseline covariates using the control group, predict $Y(0)$ for the full sample, and estimate treatment heterogeneity across groups defined by $\hat{Y}(0)$. To avoid the finite sample bias that stems from predicting and analyzing outcomes on the same sample, they use both leave-one-out (LOO) estimation and a repeated split sample (RSS) approach, though the latter performs somewhat better in simulations.

In our setting, we are interested not in heterogeneity over $Y(0)$, but rather over hours of participation, which we will call H. However, we only observe H for treatment individuals (where $H = 0$ for treatment never-takers). To define participation-intensity groups for everyone, we flip the Abadie et al. (2018) approach. Specifically, we estimate the relationship between H and baseline covariates using the treatment group, predict H for the full sample, and estimate treatment heterogeneity across two groups defined by $\hat{H}$: those with below- versus above-median predicted dosage.

As shown in Appendix Table A.XVI, observables do have some explanatory power to predict take-up among the treatment group and dosage among participants, though not a lot. Therefore, we expect this prediction exercise to be fairly noisy. Note that any heterogeneity across predicted dosage groups is not necessarily evidence of a dose-response relationship. It is possible that the kinds of people who are inclined towards high participation are more responsive to the treatment for other reasons, and still would have been more responsive had they participated for fewer hours. In other words, this exercise does not establish whether heterogeneity by predicted dosage is due to the dose itself or due to factors correlated with participation decisions.

We start by assessing how well the predicted dosage corresponds to the actual dosage among the treatment group. Appendix Figure A.IV shows the LOO predictions plotted against actual hours participated, divided into deciles with 110 people in each bin. The predictions slope upwards, suggesting that observables do capture some of the variation in decisions about how many hours to participate. But the relationship is flatter than the 45 degree line. On the right side of the plot, the actual hours participated are considerably lower than the predicted hours; observables do not do a great job at predicting who participates the most. They also under-predict at the bottom, where the lowest decile, which has a prediction near 0, actually participates just under 150 hours. In practice, perfect prediction on the left side of the graph would, in theory, show 45 percent of the sample as non-participants at 0. But our linear specification tends to flatten the predictions.

Nonetheless, as reflected at the bottom of Appendix Table A.XVII, the predictions anticipate real variation in dosage. When we divide the entire sample up into two groups by their predicted dosage, the average number of hours rises from 143 to 426 in the low- versus high-predicted dosage groups (which we will call high- and low-dosage, omitting the "predicted," for convenience). The predictions also capture variation on the extensive margin of participation, with take-up rates rising from 39 to 70 percent. As such, the variation in ITT effects between these two groups can be informative about whether people who are inclined towards higher levels of participation have a bigger treatment effect.

The table reports ITTs for each group across our primary index and its components. The pattern is suggestive, though a bit too noisy to be definitive. For the LOO estimates on the low-dosage group, the point estimates for most outcomes are positive, if often small, and not statistically significant. All LOO point estimates for the high-dosage group are negative, though none are statistically significant. The standard errors mean the two groups cannot always be distinguished from each other, but the pattern is consistent with larger effects for those inclined toward greater participation.

A similar pattern arises in the RSS estimates, where the shooting and homicide arrest result is statistically significant ($p = 0.02$) and the p-value on the index is just above 0.15. For the RSS estimates, the low-dosage group has mostly negative point estimates that are all relatively small both in overall magnitude and relative to their standard errors. The high-dosage group, however, has negative point estimates that are substantially larger than those in the low-dosage group (with the exception of other serious violent-crime arrests, where as in our main results, the point estimate is positive). For example, the high-dosage point estimate is 4.2 times larger than the low-dosage point estimate for shooting and homicide arrests. Given that the take-up rate is only 1.8 times as large for the high-dosage group, even the LATEs implied by these estimates would be considerably larger for the high-dosage group than the low-dosage one.

Overall, then, both methods for avoiding finite-sample bias in this prediction exercise suggest that participants who participate more have suggestively larger effects. This pattern cannot isolate whether high-dosage participants would respond more to the same dosage or are responding more *because of* the higher dosage. Indeed, the patterns shown in Figure III suggest that some of this difference is due to the different pathways that contribute to the low- and high-dosage groups. That figure shows that within pathway, dosage does not appear correlated with treatment effects. Consistent with that evidence, when we repeat the endogenous stratification exercise but omit pathway as a predictor of dosage, we no longer see the same pattern of results: point estimates for the low-dosage group are always negative and of comparable or greater magnitude to those in the high-dosage group. Taken together, these results suggest that the differences in predicted dosage groups may be more about who participates more—and in particular the unobservables that outreach workers are using to make referrals or their difference in program delivery—rather than the fact that they participate more. Given the expense of offering an 18-month program, future research that experimentally varied the dosage would be quite helpful.

A.7 Details on benefit-cost comparison

Assigning dollar values to the harm generated by crime is not a straightforward task. It involves a number of complicated conceptual and ethical issues, including how to value the loss of life. Our analysis is motivated by the excellent discussion of these issues in Dominguez and Raphael (2015), which emphasizes the uncertainty inherent in the estimates, along with the care with which policymakers should use benefit-cost comparisons as an input into—but certainly not the sole input into—decision-making.

We do not aim to conduct a comprehensive benefit-cost analysis. In particular, we focus on the social costs of crime, setting aside other potential benefits and costs of the program: the opportunity cost of the dollars that fund the program, the work that program staff might have done in the absence of their jobs on READI, the benefits of investing in under-served neighborhoods, the value of the work READI participants did, other unmeasured benefits of the program, and so forth. We do not mean to trivialize the potential importance of these benefits and costs. But there is little empirical work to help guide our thinking on the social value of these aspects of the program, and we have few good measures of how much they changed. As such, assigning dollar values would necessarily involve a huge amount of speculation and extrapolation.

By taking a narrower focus on the costs of crime, we can rely on a longstanding literature that estimates the social harm from victimization, as well as the cost of running the criminal legal system and punishing offenders (although even these latter costs may be still be underestimated, since the broader harms that can come from policing, criminal records, and incarceration are an active area of ongoing research). This narrower focus is not intended to generate comprehensive estimates, and has a number of real limitations. But it also provides a concrete way to start to quantify the value of programs that aim to reduce violence, which as Dominguez and Raphael (2015) argue, should be a central input into decisions about how to direct resources.

Our aim in this section is to provide transparency about how these estimates are constructed, what is and is not included, and how sensitive the estimates are to the biggest outlier in terms of social costs: the cost of a lost life.

A.7.1 Cost of crime estimates

The literature on estimating the social harm from crime typically takes one of two different approaches. The first is a "bottom-up" approach, which uses observable prices from pieces of the harm a specific crime generates—medical and legal bills, mental health and social services, rising insurance costs, jury awards for pain and suffering, and the like—and adds them up to approximate the direct costs of different kinds of victimization. The seminal estimates of this approach come from Miller (1996); we use the updated version of their estimates reported in Cohen and Piquero (2009), inflated to 2017 dollars.

The second is a "top-down" approach, which uses contingent valuation surveys to estimate a broader willingness-to-pay (WTP) to avoid different types of crime. These WTP measures in theory capture a broader set of social harms that what can be measured in the bottom-up approach, including the investments people make to avoid crime in the first place. In part for this reason, they tend to be somewhat higher than bottom-up estimates. We use the WTP measures from Cohen and Piquero (2009), inflated to 2017 dollars. Note that when people report their WTP to avoid an assault, for example, they are theoretically including the tax burden involved in arresting, trying, and punishing offenders as well as the productivity loss from incarceration. To avoid double-counting these pieces, we use the WTP measures in Cohen and Piquero (2009) after subtracting the bottom-up measures of legal costs and offender productivity loss.

There is debate about whether and how to use WTP estimates in valuing crimes that involve the loss of life, since WTP is shaped by ability-to-pay and therefore varies by factors like income. As Sunstein (2013) argues, it may be appropriate to consider an individual's own WTP for a reduced mortality risk when they are being asked to fund the policy directly. But when external organizations are funding the intervention, as is the case with READI, there is a clear rationale for treating the value of life equally across people, or even up-weighting the protection of lower-income lives based on redistributive welfare gains. As such, we follow all federal agencies in using a unitary value for the harm from a loss of life. There is still some uncertainty about the appropriate dollar figure to use, informed by considerable research outside of crime on issues like climate change and transportation policy that involve a risk of mortality (see, e.g., summary in Sunstein, 2013). We view the different dollar figures across the more and less inclusive columns in Table VII as a way to reflect some of that uncertainty.

The first 5 columns of Appendix Table A.XVIII report the dollar figures that correspond to both types of cost estimates by crime type. Note that these dollar figures represent a rough consensus from the literature, although there is certainly variation in the details (see, e.g., Chalfin, 2015). For this reason, along with the conceptual issues about what types of costs should be included, the standard errors in Table VII under-estimate the amount of true uncertainty in our estimates.

A.7.2 Details on calculations

The left panel of Table VII attempts to make a series of more conservative decisions, which likely understates the true harm from crime, while the right panel makes a series of more inclusive decisions which may come closer to estimating the true amount of harm, but are necessarily more speculative. The two sets of estimates can help give a sense for how much some of the measurement issues surrounding our outcomes might matter.

In both sets of estimates, we assign to each arrest and victimization of a READI study member an estimated cost of crime depending on its type.[13a] In the more conservative estimates, we use the lower bottom-up costs of crime. We also limit the calculation to incidents that appear as a victimization or an arrest in our data. That is, we only count a victimization as occurring if it is reported in our victimization data as having a victim in the READI study sample. For shooting and homicide victimizations, reporting rates are close to 100 percent, so this should be reasonably close to a complete measure. Lesser crimes are not as consistently reported to the police, so these counts understate the amount of actual victimization that occurs.

The more conservative estimates also only count an offense as occurring if it is reported in our arrest data with an arrestee in the READI study sample. This may overstate actual offending in cases where the person arrested did not actually commit the crime. But the magnitude of this overstatement is likely quite a bit smaller than the under-reporting stemming from the fact that most offenses are not cleared—i.e., they do not result in an arrest.

[13a] In principle, an arrest may be for an offense that resulted in multiple people being victimized, and multiple arrests may be associated with a single victim. In practice, we cannot reliably measure the number of victims associated with an arrest due to data limitations. As a result, we assign a single cost to each arrest and victimization in the data, which is standard in the cost of crime literature. While this will underestimate the social costs associated with some arrests (those with multiple victims per arrestee), it will overestimate the social costs associated with other arrests (those with multiple arrestees per victim).

Clearance rates are often quite a bit lower than reported crime rates, so on net we expect arrest counts to understate the amount of offending that actually occurs among the READI sample. One implication of only counting offenses that appear as arrests in our data is that the conservative estimates do not include the costs of the legal system for those are arrested for victimizing a READI sample member, but are not in the READI sample themselves.

The more inclusive cost of crime estimates use WTP measures of the social harm of each crime type. They also make adjustments for the under-reporting of victimizations to police and the fact that most offenses do not result in an arrest. To make these adjustments, we use the crime-specific reporting rates (call that rate R) from the Bureau of Justice Statistics (Morgan and Thompson, 2021). We calculate separate reporting rates for 2019 and 2020, then average these two values to estimate the overall reporting rate.

A complicated issue surrounds the time coverage of our calculations. When a crime is committed in our data, we use estimates of social costs that span a victim's lifetime. But due to their involvement in crime, it is at least theoretically possible that saving the life of a sample member could generate other, future costs in terms of increased offending after our 20-month outcome period. Because it is difficult to anticipate how READI will affect future behavior, we do not attempt to extrapolate beyond the crimes we observe in our 20-month outcome period. But it will be important to follow the population over time to better assess total costs and benefits.

For shootings and homicides, we use clearance rates that are specific to our context, calculated from the Chicago Police Department data. We do not have clearance rate estimates for other crimes, so we use prior work that has estimated the number of crimes that actually occur for every observed arrest (call that rate C). In particular, we use the midpoint of the range given in the Belfield et al. (2006) benefit-cost analysis of Perry Preschool, giving preference to felony rates when available. The resulting victimization and clearance rates are reported in the last two columns of Appendix Table A.XVIII.

To be clear on what enters each row of Table VII: The "READI Sample Victims" row assigns social costs to each incident where someone in the READI sample was reported as a victim in the data. The more inclusive estimates inflate each victimization by the average reporting rate for each kind of incident, and use the WTP social costs. The "READI Sample Offenders" row assigns social costs to each incident where someone in the READI sample was arrested. The more inclusive estimates inflate by both the clearance rate for each type of arrest (to account for reported offenses that do not result in an arrest), as well as the victimization reporting rates (to account for offenses by READI offenders that do not get reported to the police and so could not feasibly be cleared by an arrest).

We assign the average legal system cost and productivity loss from incarceration for each type of crime. The more conservative estimates assign those costs to each observed arrest only. The more inclusive estimates add the legal and incarceration costs of the arrests that are implied by READI victimizations. That is, they assume that when a READI sample member is victimized, someone else is arrested at a rate equal to the clearance rate for that crime type.

It is not straightforward to calculate the exact cost per participant (or per random assignee), as program operators lump fixed and variable costs together, and use budget periods that have varying numbers of participants cycling through them. In addition, the COVID-19 pandemic shifted costs around in unexpected ways. We currently only have time windowed cost information pre-COVID, so our estimate of the administrative cost of the program is only from the pre-COVID period. To approximate the administrative costs of the program, we calculate average monthly spending on everything for the program (staff, participant wages, legal and insurance costs, and other operations). We then calculate the average number of participant-months during that period, use that to calculate the average monthly participant cost, and multiply by 20 to match the costs accrued during the average 20-month period. Note that this is not quite what we would want in principle, since it includes fixed start-up costs (spread only over the pre-COVID period), and the early part of the program may not be representative of the spending for the program as a whole.

We also only have an approximation of how many of these costs went directly to participants as a transfer via wages and CBT/training stipends. For these, we use approximations from the payroll data, which have their own imperfections. Given the measurement error, we have used what we believe is a lower-bound of payments to participants so that READI's net cost appears as high as possible (making it harder to find a positive benefit-cost ratio).

A.8 Qualitative data collection and analysis

The goals of the qualitative data collection efforts for READI Chicago were to: (a) describe participants' experiences and perceptions before and after participation in the program; (b) inform ongoing and future program design; and (c) aid in the interpretation of impact estimates.

We focus our attention in this appendix on providing supporting material for claims made in the paper on several topics: assistance to the control group; the relative importance of program components; how outreach workers selected and worked with participants by pathway; and, finally, qualitative evidence on mechanisms.

We collected most qualitative data early in the program, prior to analysis of program impacts. We can and do use emergent themes from these interviews to try to understand patterns and outcomes in the quantitative data, but it is important to note that the data collection activity was not designed to interpret specific ex post findings.

A.8.1 Qualitative data sources

We have three main sources of qualitative data collected across all three READI sites: (1) approximately 220 hours of field observations both from the investigators and from a team of qualitative research assistants, (2) 16 structured focus groups conducted with 90 percent of all front-line staff in Spring 2019 ($N = 84$), and (3) 23 semi-structured interviews with program participants.

In addition, in Winter 2020, research assistants had begun piloting a survey instrument that would then be systematically administered across sites. Research assistants administered 66 pilot surveys to unique individuals in the treatment group in the Austin/West Garfield Park site in Winter 2020 before efforts on survey administration were curtailed due to the COVID-19 pandemic. The majority of surveys were administered at job orientation, though some participants had been participating in the job for several months at the time of survey administration. These are the surveys we use to report employment details in Section 4.1. Here we elaborate on the three main sources of qualitative data.

Field notes The qualitative research team used an ethnographic approach to field observations that focused on hypothesis generation around participant experiences and program design and implementation. Research staff conducted site visits with work crews, outreach workers, and employment organizations to understand and document the key ingredients of READI. Staff did not attend the group therapy sessions, which were considered private, but were often on site before and after the sessions. Many of the site visits were concentrated in the early months of READI, to get a sense for how the program was being implemented on the ground, and to get an informal sense of fidelity to the model. Researchers used "points of observation guides" to guide questions, observations, and note-taking. The primary focus was on the everyday operation of READI, what was working well or poorly, and how staff and participants perceived READI. Informal conversations with participants and staff were also a way to pursue some of the same questions used in the focus groups and participant interviews, but these results should be regarded as anecdotal. Research assistants documented their field visits in detailed, ethnographic observational notes no more than 24 hours after their visit. Details captured in these field notes include detailing sensory elements of the visit (e.g., the temperature of the room, whether it was quiet or raucous, describing the physical space in which programming occurred), describing groupings of individuals and where they were standing in a room, recapping conversations about the program and perceptions with READI staff or participants, describing programming as it occurred (e.g., how participants were loaded on the vans and transported to the worksites), when and why conflicts arose, how they were resolved, and other such situations of note. Field notes provided rich narrative text to understand what was happening on the ground across multiple sites as the program was implemented over time.

Focus groups Focus groups followed structured protocols and asked participating staff, grouped by role (outreach workers, coaches, work crew chiefs, managers) and by site, to comment on the following subjects:

1. Barriers to program implementation, including what READI does well or poorly
2. The types of participants for whom they think READI works best or worst
3. Descriptions of participants who have left READI or refused to take part in the program
4. Which READI program components they think are most important and why
5. How assignment to the control group was communicated to outreach referrals, and whether the staff are aware of what the control group subsequently did or any programming they received

Focus group facilitators took detailed notes on the discussion and answers of most participants, and asked for voting on some key questions such as which READI program component was the most important. In addition, focus groups were recorded and transcribed using an automated transcription service, Rev.com. Due to constraints in audio quality, research assistants reviewed all focus group transcripts as part of a quality assurance process and used detailed notes to supplement transcriptions where necessary.

Participant interviews Two qualitative research assistants interviewed 23 participants beginning in Spring 2019, approximately 18 months after READI launched and when implementing organizations felt comfortable allowing research assistants to work directly with participants. Participants were selected based on a snowball sampling method where research assistants would ask participants to participate during their field visits and/or through referrals of willing participants by outreach workers or work crew chiefs. Given high levels of initial distrust among participants, interviews were conducted in the first months of their participation, after sufficient trust had been established. These interviews were only semi-structured, and were first and foremost focused on the subject's personal narrative—their background, why they were interested in READI, their initial experience in the program, how it was affecting their relationships (including with adversaries), and what successes, barriers or difficulties they were facing. As a consequence of the timing and focus of the interviews and sampling method, these interviews do not speak systematically to program mechanisms or long-run impacts, or even causes of violence. Some of these themes emerged from the interviews, however, and should be regarded as generating hypotheses for future research. Interviews were recorded and transcribed with participants' consent.

A.8.2 Methodology

Our main approach to qualitative analysis is the use of emergent themes (Williams, 2008). In this style of analysis, patterns are not predetermined or imposed by the researcher, but rather emerge from the data during repeated readings. This approach is commonly used in exploratory research. Specifically, the researcher reads through the data multiple times to identify and code keywords, patterns, and commonalities. These patterns are then grouped together to form themes. The themes are continually refined and re-evaluated as new data are collected.

This approach tries to allow themes to emerge naturally rather than being influenced by preconceptions or biases. It can also help reveal unexpected insights and perspectives that may have been missed with a more structured or predefined approach to analysis.

Emergent themes can be coded and quantified, to a degree. Sometimes it is appropriate to report this in proportional terms or as strict rankings. With our focus groups, for example, we asked most participants the same structured questions, and thus we can report these results in percentage form. When it comes to the field observations and participant interviews, however, it is generally considered inappropriate to translate emergent themes into strict rankings or percentages, most of all because of the data generating process. (This is especially true when it comes to unexpected results, such as the differences in impacts by pathway, or the importance of unobservables to outreach worker recruitment. Because we did not systematically ask about these topics in interviews, our emergent data are less systematic and quantifiable.) Therefore, when it comes to analyzing participant interviews and field observations, our analysis will generally seek to provide a number of representative and illustrative quotations, as well as a quantitative sense of their frequency.

To code keywords and themes, we used the software program NVivo. Research assistants transcribed, read, and coded over 3,000 pages of recordings and field notes using an emergent or inductive coding scheme. First, the qualitative research team worked with the authors to create a base coding structure. Next, the team systematically read and coded the raw text into short segments of text. Text segments were grouped together and categorized based on the initial coding scheme but allowed for emergent themes. Excerpts could be double-coded, but not triple-coded. The resulting coding structure was periodically updated based on ongoing analysis, ongoing observations from the field, and frequent discussions among the team about emergent themes.

This coding process generated 9,235 uniquely coded excerpts across 338 categories. Research assistants then analyzed the categories to determine the most frequently cited responses to questions, as well as major themes and subcategories that emerged from the text. The top-level categories and themes were then synthesized and supported by quotes from across data types to represent the meaning of the category or theme.

All items were coded in NVivo to track inter-coder reliability. Research assistants double-coded 10 percent of raw text to ensure inter-coder reliability, for a rate of 91 percent agreement. To ensure that the coding team agreed on the coding structure, all coders coded the same document and met weekly to revise the coding structure and create shared definitions. In addition to the coding structure, the team created a guiding document that defined codes, listed criteria for inclusion or exclusion, as well as example excerpts.

A.8.3 Control group activities

A natural concern is that control group members may have received other programs, biasing treatment effects downwards, and so this was an explicit focus of our qualitative research plan. The data suggest this bias is present but may be modest.

Our qualitative data on this topic come primarily from field observations. One of the six key priorities for this activity related to the control group. Field observers were instructed to look for resources that outreach staff provide to the control group post-randomization, and specifically if there is any indication they are referred to non-READI programs. In addition, some comments related to control group engagement emerge from the staff focus group interviews, as outreach workers were each asked, "Can you tell me what some of the [outreach referrals] that didn't get in are doing now?" (There are no data from participant interviews, since no control subjects were interviewed.) Drawing on our systematic coding, we reviewed all references to control group engagement.

There is evidence that outreach workers attempted to connect READI control group members to other social services and job opportunities. In at least five coded excerpts, outreach workers reported attempting to refer the outreach referrals assigned to the control group to other programs run by other non-profits. For example, one explained:

> I say it's a 50/50 chance [to get into READI]. Then, people who are in the control group I ask them, "do you have a license, and can you drop clean?" Then I connect them with other services.

One field observation noted:

> [Outreach worker name redacted] said that they interact with the [outreach referrals] randomized into the control group on a daily basis. Some of these guys

> are participating in other their organization's other programs, such as the detention reduction program, legal intake, neighborhood interactions, or another CBT program.

Outreach workers reported that they and their organizations still had obligations to the control group members. According to one outreach worker: "if we see a [control group] guy that was a shooter, we still have to help them anyways." According to another outreach worker:

> You went [said] to control, "you still got an outreach worker. You still got somebody that's gotta help you. Go down there and make their ass work, don't let them go downstairs and sit there and do nothing. You didn't get in, that don't mean we not gonna help you." Cause when they go control, they still got an outreach worker downstairs that's still gonna work with them, so I tell them.

It is unclear based on the qualitative data what the frequency or success of these referrals were, or the length of subsequent engagement. Overall, however, in repeated field visits by the investigators and qualitative researchers, this engagement of the control group seemed limited. Outreach workers had a huge portfolio of participants to track down. Helping existing participants manage their many daily difficulties was also time consuming, and so the ability of outreach workers to assist the control group was modest.

When outreach organizations did refer control group members to other services, it is also worth noting that almost none of the alternative programs available in these neighborhoods offered programming as lengthy or as intense as READI, and almost none offered sustained paid work, or work with CBT. As one field observer mentioned:

> I asked the three [outreach staff] if they had interactions with the control group. They all said not really, except for when they ran into them in the community. I then asked if there were any [their NGO] or local programs that they offered those not READI? [Name redacted] said there is nothing like READI around here. She explained that many are disappointed they didn't get in, and she hopes READI can serve them eventually once the study is over.

A.8.4 Perceived importance of program components

Another explicit focus of the qualitative research was the relative importance of the various treatment components, and here we have some of the most complete data. Generally speaking, most staff agree that all components were important, with the job and paycheck being a particularly important incentive to participate and a place to practice skills learned in therapy, but the skills learned in therapy themselves as the foundation for behavior change.

The most systematic data on the importance of program components come from our focus group interviews with staff. We first asked about the three pillars of READI: CBT, the "relentless engagement" of outreach workers, and the job. We asked staff to vote on which was the most important program component and explain why. We also asked, when it came to the job, which was more important: going to work or getting paid? The results below gather conclusions from the two related topics of discussion.

Appendix Table A.XIX summarizes voting responses, first for all focus group participants and then separately by the role of the staff person. In all cases there was disagreement. 19 percent of staff suggested that the therapy was the most important component; 15 percent cited relentless engagement; and 11 percent pointed to the job. But the balance (35 percent) resisted nominating a single component and emphasized that it was a combination of components.

In favor of the job component, those who conducted the focus groups summarized that most speakers agreed that participants' main motivation for working was to receive a paycheck. They speculated that this money was needed to support participants' families or to fund drug habits. "Let me answer it like this," one staff member explained, "I don't think they would come to work if they weren't getting paid at the end of the week." Indeed, many staff agreed that receiving a paycheck was the main motivating factor in participants' attendance.

Nonetheless, twice as many staff emphasized CBT as the most important component for behavior change over the job and payment. Many participants were originally motivated by the money, they explained, but experienced a turning point where the CBT curriculum and other aspects of the program became just as, if not more, valuable. Participants realized that they could apply CBT to other aspects of their lives and use their work experience as a stepping stone to better things. In addition to receiving a paycheck, speakers also cited the atmosphere of READI as a motivation for participants to attend. They believed participants saw READI as a safe place where they could avoid the violence in their communities and build relationships with the program's mentors and one another.

The largest proportion saw the combination of these approaches as essential. To give an example, even staff who thought the therapy was the most important source of behavior change sometimes conceded that many participants are motivated to come for the money. In addition, several acknowledged that without the constant efforts to keep the participants attending and engaged and feeling supported (relentless engagement of outreach), the therapy could not be effective.

Finally, in addition to these specific program components, some outreach workers spoke about the incapacitation effects of a program like READI that had men off the streets for both CBT sessions and work: “We get them out of the streets,” one said, continuing: “Like, because for program shut down days, like sometimes we either leave early or like even if we didn’t work that day... I live here in N Lawndale, I’m driving around, I see them. I see them in the corners; I see them doing stuff they’re not supposed to be doing. So I know like when they’re here at least they’re not doing that.” It is important to note, however, that we do not find empirical support for this hypothesis in the quantitative data (see Appendix A.5.6).

A.8.5 Selection and recruitment by outreach workers

Another explicit focus of the qualitative research was how outreach workers recruited men into the program. This is largely because no program had previously attempted to recruit men facing such high risk, and none had attempted to do so “cold” as with algorithm referrals. Moreover, we anticipated imperfect compliance and selection into the program, and we wished to document possible sources of selection and how this could affect program impacts. Later, after learning about the heterogeneity in impacts by pathway, the subject of recruitment and differential patterns of selection took on greater importance.

In brief, outreach referrals often featured higher levels of personal familiarity, and they were screened to find men with very high $\hat{Y}(0)$ who at the same time were personally committed to changing their lives, did not have any severe levels of mental health and substance abuse challenges to preclude program participation, and who were not already on the path to change or for whom other programs had not seemed to work to support behavior change. This type of screening can help to explain the higher take-up rate of outreach referrals, but it could also result in more precise and arguably larger treatment effects for this group compared to algorithm referrals who were not similarly screened.

The role of personal familiarity, relationships, and networks Outreach workers typically had close personal relationships with their referrals. Sometimes these relationships spanned several years or generations, and other times they were formed during the recruitment and screening process. These relationships were explicitly mentioned in at least three interviews, although their depth and quality was self-evident in innumerable informal observations. One outreach worker said, “Usually these are guys we already know, probably for years.” Another, referring to a recent recruit, explained, “I been knowing him since he was a baby.” And one participant explained in an interview his personal connection to an outreach worker:

> She knows my mother. My grandmother said she’s from that area. Her mother lives like literally down the street. The street that’s in front of my grandmother’s big window. Her mother stays right down the street on the next block and then her mother and my mother go to the same beauty salon. So they talk all day,

just gossip.

In other cases, outreach workers did not have a preexisting relationship, but rather received a referral through their information gathering. Even in these situations, however, outreach workers tried to get to know the individual well in order to assess their risk factors and suitability for READI over repeated visits. One outreach worker commented that he looks for:

> ...how they engage you, the questions they ask you. That's how you know how interested they are. Some days when I talk to them, you know, I see how they live. Like I go to their house, and usually a female answer the door—which more likely, it's her house. Then when I talk to him—you know, explain stuff to him—I try to see if they intake what I'm saying instead of everything I say they questioning me. [Also] are they high? Most of them you go to their houses, you come in it smell like weed.

One outreach supervisor explained that this personal connection is essential. "I get very upset with people who put in [an outreach referral] and then can't even connect with them in the end," he explained, "Why did you put them in here if you didn't think it would work?"

One possibility is that these close relationships and information could have made the program more effective for outreach referrals—a direct treatment effect rather than a selection effect. For instance, some staff discussed how preexisting relationships may have enabled or motivated staff to provide more/better support for outreach referral participants by understanding more about their backgrounds and personal obstacles to provide the right supports, or by being partial towards them. As one outreach worker explained, "I been knowing him since he was a baby, you know, his family we like this. So (laughs) that's like my second home. And, for me to get him in the program and him not to show up, and not come—you know it's like, I'm really hot with him."

These preexisting relationships may also have made the outreach referral participants feel more motivated to show up to READI, out of personal gratitude or obligation. As one participant explained about outreach workers: "they'll buy you a meal, they'll do this, they'll pick you up for work. You know, they show you they care. So why would you come here and like, you know, make them look bad."

Another theme that was surfaced in our data was that outreach workers sometimes recruited men in networks or in groups. It is conceivable that, by joining the program with friends, they were primed for READI, and more likely to feel safe and continue to attend despite coming up against unfamiliar situations, people, or rivals. One outreach worker mentioned that he noticed that if the most senior gang leader attended, lower level gang members were more likely to attend, and vice versa. In principle, this social network effect for take-up and participation may also have influenced receptiveness to programming and program efficacy.

The chief reason for skepticism of a direct treatment effect of personal relationships is that dosage data show that outreach referrals were no more likely to remain in the program than algorithm referrals at the same level of predicted risk. Thus, the effect of these personal relationships and information on dosage and efficacy may be less important than the effects on selection and take-up. The remainder of this section explores these selection and take-up effects.

Efforts to select on risk of gun violence involvement in the outreach pathway Outreach workers were first and foremost instructed to refer men based on their risk of involvement in gun violence ($\hat{Y}(0)$). This was a mantra of the program. Most of their training and supervision was focused on identifying the highest risk cases—especially men engaged in disputes or active rivalry. "We trying to save some lives," was a familiar refrain.

How did they identify such men outside of the algorithm? Outreach workers typically received this information through their social networks, active canvassing and inquiries on the street, or by observing social media. Outreach teams and supervisors regularly met and discussed each candidate's risk profile before submitting them to the research team for randomization. Although the research team did not assess each individual outreach referral's risk score, we did share information on the aggregate risk scores of outreach referrals with Heartland Alliance to assist them in making sure that READI was reaching its target audience (see Appendix A.3.1).

Finally, in some cases, outreach workers were explicitly recruiting subjects from opposing groups ("opps"). By using social media and personal contacts, outreach workers were actively monitoring rivalries. Once identified, outreach workers might need to do considerable mediation or preparation between the two rivals before bringing them into the same room, for safety reasons. This could take days or weeks and was also an opportunity to build trust and get a better sense of readiness for the program.

At the same time, outreach workers were trying to balance propensity for violence against willingness to engage and the danger involved in engaging an uncooperative or volatile participant. We discuss this selection on "readiness" below. Still, it is worth noting that outreach workers generally tried to engage and help "unready" men to eventually get them into the program (regardless of referral pathway). "You might have to take more time with them," one outreach worker explained, "because your name is on the line and also this person get in the program could be a tyrant and [colleagues will say], 'Oh man you knew what type of person he was.' "

Screening by responsiveness to treatment Finally, field observations also focused on the extent to which outreach workers were trying to screen candidates for responsiveness to the treatment, and screen out those whom they believed would be unlikely to respond to treatment. Exclusion criteria mentioned include: those with serious mental health conditions, serious drug abuse problems, those who already had jobs or educational opportunities, and (most frequently of all) those uninterested in changing in their lives. As one outreach supervisor explained:

> ...[W]e can't afford to lose them. I tell outreach workers don't put someone here you can't find and who doesn't want to work. Why the fuck you put his name in if he don't want to work. You wasted a slot, you could have saved somebody else's life. That's crazy, we have to be serious about the job and community referrals. Guys who fit this criteria can benefit from it. They need to make it click that this is what READI is going to get you.

By far the most common theme that emerged in focus group and field interviews was that outreach workers were screening out participants who were not "ready for READI." One outreach worker said: "I wouldn't say that any of them are unfit for READI. Um, I think it's just, their level of... willingness to change their lives." This theme arose in at least 11 focus groups and more than 25 of 346 coded narrative excerpts, including some of the quotes above.

Some of these screenings are formal through an interview process, while other are informal based on the outreach worker's discretion. For example, one outreach worker explained:

> When we see the guys that's ready for change, we have to give them more attention. Not saying that they don't deserve more attention, but here, he ready. So, why let him go back to the flock? ... So, I really gravitate to him and-and and try to get him more, because, you know, working with this job, we wanna save everybody. You want everybody to get it. But everybody's is just not gonna get it. And it's gonna take some of these guys one, two, three or four, five, or six times before they get it. But the ones that come in and ready—them the ones, you know, them the ones who, collectively, all of us as group try to push.

Another outreach worker said he looked for signs of personal commitment in subjective, informal assessments of truthfulness:

> you have to fit a certain criteria, it's just not for everybody so it's just can't submit everyone. So I need them to be as truthful as possible. I know a lot of guys like to hide things so I make sure they're real comfortable, expose all your

> answers, let me know what's going on with you, and then maybe you can go in the lottery with your name, maybe get picked.

Almost every outreach worker described a different technique they used to assess readiness. Above, we described the efforts to develop personal relationships, visit men at home, and get to know people in their lives. Others look for costly signals of seriousness. One said that his organization had the candidate come in multiple times to the office to signal their seriousness. In another case, the field observer noticed that a candidate outreach referral who came into the office was later screened out of eligibility for READI by his lack of engagement and willingness to share his identity documents—a clear lack of trust.

Sometimes, outreach staff explicitly communicate this criterion to candidates. As one outreach worker explained:

> We tell them like, don't take this spot if you ain't serious. So many people want it, coming for it, and I be like: "Man, if you really finna [fixing to] work and you really finna to do it. If you not, don't take this chance, because they only get so many guys." So, the ones that I put in I make sure they serious about coming to work, and I let them know.

Participants echoed this sentiment. In one exchange:

> Interviewer: Is there a particular kind of person that you think READI works best for?
>
> Participant: A person willing. Because I mean they give you all the tools, they give you everything. You just got to be willing to want it. You got to want it. It don't matter what type of person you is. If you're a person that like money, if you are a person that like talking, if you're a person that like any, I mean they got something for you here, whether it's a job, whether it's a train, whether it's an ear to listen to. I mean, they kind of meet everybody, long as you want it. Long as you want to better yourself.

Of course, there were several other exclusion criteria in addition to a lack of "readiness". For example, in five focus groups, outreach workers explained that they also screened on serious mental health issues and drugs, which might account for positive selection into the program. As one explained:

> When referring 'em, I look for substance abuse issues, mental health issues, basic health, basic knowledge, general understanding. Just to see, basically, are they ready for this because you can be willing, but you could not be ready. So we

> have to make sure they're ready to turn over to TJ [transitional jobs] before we do things that we do.

For example, when it comes to drugs, recall the outreach worker quoted above:

> Some days when I talk to them, you know, I see how they live. Like I go to their house, and usually a female answer the door—which more likely, it's her house. Then when I talk to him—you know, explain stuff to him—I try to see if they intake what I'm saying instead of everything I say they questioning me. [Also] are they high? Most of them you go to their houses, you come in it smell like weed.

Finally, outreach workers also screened out men who appeared to be on a path to change, such as those who had jobs or were enrolled in school, which might result in negative selection into the program. For example, four outreach staff and their supervisors said that, prior to randomization, they attempted to screen out candidates if they had another job or if they were enrolled in full-time school (e.g., a GED program).

A.8.6 Mechanisms

Over the course of data collection, we designed our observations and interview guidelines to elicit evidence on the following mechanisms:(1) incapacitation and change in time-use, (2) meta-cognitive benefits that affect participant behavior(s), (3) increased financial opportunity due to participation in the formal labor market, (4) differential social networks, and (5)differential interactions with law enforcement. In addition to our five ex ante hypothesized mechanisms, we identified two emergent themes that suggest alternate mechanisms for treatment effects: (6) increased future orientation and hope, and (7) increased social costs for risky behavior because of strengthened personal relationships and changing sense of identity. In total, there were 397 coded excerpts related to mechanisms out of our total 9,235 coded excerpts.

Incapacitation and change in time-use We found some evidence for the hypothesis that READI kept men occupied during times that they might otherwise be engaged in illicit activities. 11 of 397 coded excerpts referenced incapacitation.

For example, some participants mentioned staying at home on nights and weekends so as not to be tempted to fall into old habits. Others said that the length of time of READI programming meant they were tired on the weekends and did not want to engage in social activities. It is possible that this kind of differential time use and/or changing social networks is one avenue for behavior change, but coupled with evidence from our quantitative analysis (see Appendix A.5.6), this does not seem like the dominant explanation for treatment effects.

Meta-cognitive benefits that affect participant behavior(s) The qualitative research elicited evidence consistent with the hypothesis that improved meta-cognition through CAD changed behavior ($N = 74$ coded excerpts). In particular, many participants and staff members mentioned the benefits of CAD programming in supporting behavior change. In the words of one participant:

> Participant: I liked to, you- you know, uh, talk about your problems, you know? Ways you can resolve, um, problems from not going to negative, like positivity, counseling, you know, showing you another way, basically. You know?
>
> Interviewer: Do you feel like you learned new things in there?
>
> Participant: Yeah, how to handle my anger. And doing something like, probably, somebody making me mad was the thing, saying on my way, how could I go about it in a positive manner? So, I know how to cope with that better. You know?

Increased financial opportunity due to participation in the formal labor market The qualitative research also elicited evidence consistent with the hypothesis that financial alleviation supported participants' behavior change ($N = 11$ coded excerpts). Several staff members discussed participants' financial status in focus groups. One staff person commented that the program is like an iceberg, where the surface level is the gift card payment, but there is much more underneath. Many staff members mentioned that participants come to the program just for the check, but as they progress through the program, they begin to see the value in it. They can use the skills they learn in the program to advance their career and gain more opportunities, such as forklift certification and going to school.

Differential social networks In several participant interviews, participants mentioned changing who they spent time with and how they interacted with their acquaintances and friends who were still involved in the informal labor market:

> Interviewer: Have your affiliations changed?
>
> Participant: Yeah, most definitely. Like... And- and you mean about affiliations just based off do I still socialize myself with a gang? [Yeah]. I'll say I conduct myself with them. It ain't like I stand on the block with them. It's like if I- I have a- an interaction with them, it - it'd be that... It's not like that I gang bang anymore, [but] it's like it's still the brotherhood that I [associate with].
>
> I'm not going to push you and say, "Hey, get the fuck away from me." I'm going to let you, who are toxic to me, [your] toxicness and let yourself drain away from [me]. Because if you know that you be having good intentions for me and that

> you know that you care, you'll stick around and you'll push for better instead of just be, "Oh no, hey, if you ain't doing what we doing no more, you ain't with it, if you don't ..." or if you ain't this that and the other and you want to be upset with me, that means you not good for me, so that means you're going to eventually drain yourself away from me on your own.
>
> And I'm not... I'm going to allow that to happen. I'm not going to fight for it.

In addition, one staff member mentioned that READI might create opportunities to create social networks across opposition, which may reduce conflict:

> READI creates a safe space. And participants have said that, you know, we have, we have seen these guys on the street and especially if their opposition that there are opps, you wouldn't even talk to these guys. But in READI, you know, bonds are formed within READI that wouldn't otherwise be formed on the streets.

Differential interactions with law enforcement Relatively little data emerged in our interviews and observations for READI's effects operating through differential interactions with law enforcement. Only six narrative excerpts were coded as relating to differential interactions with law enforcement out of 397 uniquely coded narrative excerpts related to mechanisms. All six excerpts were from participant interviews in which interviewers explicitly asked about interactions with law enforcement. Three of the six responses highlighted no change in their experiences with law enforcement after involvement in READI.

Increased future orientation and hope Future orientation was an emergent theme that was not hypothesized ex ante, but was mentioned in 50 coded excerpts related to mechanisms. Many staff expressed a sense of hope due to witnessing behavior changes. For example, participants who had a history of substance abuse became sober and started attending regularly. Staff mentioned that seeing "real and tangible" progress motivated participants—and sometimes their peers—to continue program engagement and behavior change.

Increased social costs for risky behavior because of strengthened personal relationships and changing sense of identity Much of the discussion around mechanisms with participants focused on their relationships with other participants and staff, as well as improved relationships with their own family members due to newfound communication skills. Interviewers asked participants "What made you decide to join READI?" and "What do you think of [the program] so far?" In addition, interviewers asked participants: "How has joining READI affected your relationships with your friends and family?" Eight of 23 participants spoke about how READI helped with communication with their families, specifically with women in their lives: mothers, grandmothers, and partners. Four participants spoke about how READI provided an opportunity to feel more independent ("I have my own place now") or to contribute financially for their family ("It feels good that I have $50 or $100 to spare for my grandma"). Other participants talked about feeling safe in READI, making friends with the fellow participants in their crew on the vans that transported them from the CBT sessions to worksites, and the care that staff showed towards them as important features of the program. These strengthened informal channels for social control and supported prosocial behavior changes in participants.

In sum, these emergent themes highlight the role of improved meta-cognitive behavioral skills, changing social networks, increased future orientation, increased costs to engaging in illicit behavior due to stronger relationships and changing identity and social norms. However, because of the design of the qualitative research data collection efforts, these themes should be considered emergent and not confirmatory. Future research can build on this work by collecting more systematic data on these mechanisms across treatment and control conditions.

A.9 Appendix Tables and Figures

Table A.I: Baseline balance within pathways

	Algorithm Pathway			Outreach Pathway			Re-entry Pathway		
	Control Mean	Treatment Mean	Pairwise p-value	Control Mean	Treatment Mean	Pairwise p-value	Control Mean	Treatment Mean	Pairwise p-value
N	616	616		438	440		178	168	
Demographics									
Black	0.964	0.963	0.882	0.986	0.986	0.972	0.943	0.963	0.454
Age	24.8	24.4	0.146	25.6	25.6	0.889	26.1	26.7	0.533
Primary Outcome Components, Counts									
Shooting Victimizations	0.667	0.594	0.113	0.274	0.300	0.487	0.331	0.214	0.037
Shooting & Homicide Arrests	0.086	0.073	0.415	0.068	0.059	0.559	0.079	0.107	0.346
Other Serious Violent-Crime Arrests	1.091	0.964	0.079	0.640	0.694	0.456	0.982	1.049	0.738
Risk Prediction									
Predicted Involvement in a Violent Gun Crime (Risk Score)	0.137	0.137	0.884	0.089	0.089	0.909	0.080	0.079	0.748
Missing Risk Score	0	0	-	0.212	0.155	0.026	0.225	0.179	0.157
Arrest Counts									
All Arrests	20.1	20.7	0.419	13.2	14.0	0.253	15.9	16.4	0.890
Less Serious Violent-Crime Arrests	1.9	2.0	0.557	1.103	1.102	0.982	1.4	1.2	0.202
Drug Crime Arrests	5.5	6.0	0.097	4.0	4.3	0.423	4.6	5.0	0.580
Property Crime Arrests	1.8	1.8	0.996	1.2	1.4	0.398	1.7	1.7	0.979
Other Crime Arrests	9.7	9.8	0.718	6.1	6.5	0.281	7.0	7.2	0.934
Victimization Counts									
All Victimizations	4.7	4.4	0.221	2.3	2.3	0.988	2.3	2.3	0.929
Other Violent Victimizations	3.3	3.1	0.355	1.6	1.6	0.929	1.5	1.6	0.629
Non-Violent Victimizations	0.744	0.711	0.673	0.406	0.393	0.809	0.449	0.446	0.968
Incarceration Measures									
Days Incarcerated	135.0	123.7	0.309	157.1	153.6	0.820	373.3	413.5	0.261
Incarcerated at Randomization	0.052	0.049	0.798	0.037	0.023	0.209	0.022	0.048	0.164
Joint Test									
p-value on F-test			0.120			0.825			0.131

Notes: Pairwise p-value from test of treatment-control difference using heteroskedasticity-robust standard errors and controlling for randomization strata fixed effects. Arrest and victimization measures include all available CPD data from 1999 (2010 for shooting victimizations) through the time of randomization, with counts winsorized at the top 99^{th} percentile. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. Less serious violent-crime arrests include non-Part I violent-crime arrests, such as simple assault and battery and domestic violence. Other (non-shooting) violent victimizations include sexual assault, robbery, aggravated assault and battery, and simple assault and battery. Non-violent victimizations include all other incidents such as burglary, stalking, and threats. Risk score is missing for 231 individuals who did not have at least one arrest or two victimizations within the 50 months prior to randomization (non-missing risk score Ns for algorithm pathway = 1,232, for outreach pathway = 717, and for re-entry pathway = 276). Race is missing for 38 individuals (non-missing race Ns for algorithm pathway = 1,232, for outreach pathway = 847, and for re-entry pathway = 339). Joint tests include randomization strata fixed effects and the covariates listed here, excluding all arrests and all victimizations since they are linear combinations of other variables.

Table A.II: Mean characteristics of participation prior to the pandemic

	Days to Take-Up	Hours Worked	Hours Possible	Percent
All Participants	87	457	1549	29.5%
Algorithm	137	449	1593	28.2%
Outreach	67	535	1739	30.8%
Re-entry	40	269	948	28.3%

Notes: Table shows the average number of hours participants from each pathway worked during the 20-month outcome period and prior to the start of the COVID-19 pandemic. Possible hours worked are calculated for each participant based on the number of weeks from take-up to the earliest of the end of the 20-month outcome period, 18 months post take-up, or March 15, 2020, multiplied by the maximum possible number of hours worked of 29.5/week. Variation in possible hours worked arises from differences in how long it took each participant to take up the program (shown in the "Days to Take-Up" column) and differences in when participants were randomized relative to the start of the pandemic.

Table A.III: Participants' earnings and hours, by pathway and time period relative to the pandemic

	Work		Trainings		CBT/PD		Total	
	Earnings	Hours	Earnings	Hours	Earnings	Hours	Earnings	Hours
All Participants								
Pre-COVID	$5,659	457	$121	8	$1,695	102	$7,475	567
Standby Pay	$1,211	90	$1	0	$333	20	$1,545	110
Post-COVID	$172	12	$247	16	$213	13	$631	41
Algorithm								
Pre-COVID	$5,576	449	$99	7	$1,667	100	$7,342	555
Standby Pay	$842	63	$1	0	$278	17	$1,121	80
Post-COVID	$76	5	$139	9	$160	10	$374	24
Outreach								
Pre-COVID	$6,567	535	$162	11	$1,972	118	$8,701	665
Standby Pay	$935	70	$1	0	$298	18	$1,235	88
Post-COVID	$137	10	$219	15	$173	10	$529	35
Re-entry								
Pre-COVID	$3,436	269	$60	4	$947	57	$4,443	329
Standby Pay	$2,736	205	$1	0	$545	33	$3,281	237
Post-COVID	$468	33	$553	37	$437	26	$1,459	96

Notes: Take-up defined as attending the first day of READI orientation. Hours and earnings are averages per participant over the 20 months post-randomization, within the indicated period of calendar time. Pre-COVID period includes wages and hours worked from 10/06/17 to 3/15/20 ($\sim$127 weeks). Standby pay period includes wages from 3/16/20 to 8/10/20 (21 weeks), and was calculated based on each participant's average weekly wages from the month prior to COVID. Post-COVID period includes wages from 8/11/20 to 10/08/21 ($\sim$60 weeks), when participants were given the option to return to in-person work or complete online trainings and professional development (PD), which most participants opted to do (see Appendix A.5.3). Work hours correspond to time spent at a worksite. CBT/training hours correspond to time spent in group CBT sessions, PD sessions, and online trainings. Hours and earnings are limited to men who took up and appear in the payroll data, which excludes 124 men who either took up prior to consent forms allowing the release of their payroll data being distributed or who attended orientation but did not start work within 20 months of randomization. Participation records for CBT and training before April 2020 are incomplete; hours shown are extrapolated based on available data. For additional details on hours and earnings data, see Appendix A.2.4.

Table A.IV: Robustness to alternative specifications

	CM	ITT/AME	P-value
OLS: Index pooling arrests into a single component			
Primary Index of Serious Violence	0	-0.0248	0.385
		(0.0286)	
OLS: No covariates except randomization blocks			
Primary Index of Serious Violence	0	-0.0242	0.303
		(0.0235)	
Shooting & Homicide Victimizations	0.1104	-0.0060	0.656
		(0.0134)	
Shooting & Homicide Arrests	0.0268	-0.0113	0.056
		(0.0059)	
Other Serious Violent-Crime Arrests	0.0544	0.0032	0.753
		(0.0101)	
OLS: Covariates selected using double LASSO			
Primary Index of Serious Violence	0	-0.0281	0.221
		(0.0229)	
Shooting & Homicide Victimizations	0.1104	-0.0060	0.645
		(0.0130)	
Shooting & Homicide Arrests	0.0268	-0.0113	0.049
		(0.0057)	
Other Serious Violent-Crime Arrests	0.0544	0.0032	0.746
		(0.0098)	
OLS: Alternative covariates + IPW			
Primary Index of Serious Violence	0	-0.0253	0.275
		(0.0232)	
Shooting & Homicide Victimizations	0.1104	-0.0073	0.583
		(0.0133)	
Shooting & Homicide Arrests	0.0268	-0.0108	0.068
		(0.0059)	
Other Serious Violent-Crime Arrests	0.0544	0.0026	0.799
		(0.0100)	
Poisson: Standard covariates			
Shooting & Homicide Victimizations	0.1104	-0.0058	0.663
		(0.0133)	
Shooting & Homicide Arrests	0.0268	-0.0116	0.069
		(0.0064)	
Other Serious Violent-Crime Arrests	0.0544	0.0010	0.923
		(0.0099)	

Notes: N = 2,456. First panel shows results for a version of the primary index that pools arrests for shootings and homicides and other serious violent-crime arrests into a single component, as described in our pre-analysis plan. Second panel shows main outcome results without covariates, other than the randomization strata fixed effects. Third panel shows main outcome results with covariates selected using the post-double selection LASSO (Belloni et al., 2014a,b), first partialling out the randomization strata fixed effects. Fourth panel shows main outcome results using a specification that drops several covariates we neglected to include in our pre-analysis plan, enters them as raw counts instead of indicators for groups, and uses inverse probability weights instead of strata fixed effects (which we pre-specified as a back-up plan when we thought treatment probability would not vary across strata). Fifth panel shows average marginal effects from Poisson regressions with standard covariates but excluding randomization strata fixed effects for convergence. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. CM shows control means; ITT/AME shows intent-to-treat estimates (first through third panels) or average marginal effects from Poisson models (last panel). All regressions estimate heteroskedasticity-robust standard errors. For additional details, see Appendix A.5.1.

Table A.V: Robustness to randomization inference

	ITT P-value	
	Heteroskedasticity-robust	Randomization inference
Full sample		
Primary Index of Serious Violence	0.257	0.252
Shooting & Homicide Victimizations	0.593	0.577
Shooting & Homicide Arrests	0.045	0.040
Other Serious Violent-Crime Arrests	0.733	0.740
Algorithm pathway		
Primary Index of Serious Violence	0.298	0.268
Shooting & Homicide Victimizations	0.243	0.214
Shooting & Homicide Arrests	0.572	0.566
Other Serious Violent-Crime Arrests	0.277	0.262
Outreach pathway		
Primary Index of Serious Violence	0.008	0.009
Shooting & Homicide Victimizations	0.040	0.043
Shooting & Homicide Arrests	0.009	0.007
Other Serious Violent-Crime Arrests	0.783	0.794
Re-entry pathway		
Primary Index of Serious Violence	0.312	0.359
Shooting & Homicide Victimizations	0.546	0.574
Shooting & Homicide Arrests	0.869	0.844
Other Serious Violent-Crime Arrests	0.204	0.268

Notes: N = 2,456. First column reports heteroskedasticity-robust ITT p-values on the primary index and its components for the full sample (corresponding to estimates in Table IV) and by pathway (corresponding to estimates in Table V). Second column reports p-values obtained from randomization inference that re-randomizes within strata 5,000 times.

Table A.VI: READI's estimated effects using rates to adjust for incapacitation

		Estimates				P-values
	N	CM	ITT	CCM	TOT	Observed ITT
Shooting & Homicide Victimizations						
Count	2456	0.1104	-0.0072	0.1105	-0.0132	0.593
			(0.0134)		(0.0237)	
Rate	2436	0.2804	-0.1421	0.3737	-0.2597	0.030
			(0.0655)		(0.1152)	
Rate, winsorizing outliers (99.5th pctl)	2436	0.1852	-0.0509	0.2071	-0.0930	0.030
			(0.0235)		(0.0412)	
Rate, winsorizing outliers (99th pctl)	2436	0.1761	-0.0434	0.1931	-0.0793	0.048
			(0.0220)		(0.0386)	
Shooting & Homicide Arrests						
Count	2456	0.0268	-0.0120	0.0340	-0.0220	0.045
			(0.0060)		(0.0106)	
Rate	2436	0.0459	-0.0132	0.0519	-0.0242	0.344
			(0.0140)		(0.0246)	
Rate, winsorizing outliers (99.5th pctl)	2436	0.0429	-0.0162	0.0525	-0.0297	0.157
			(0.0115)		(0.0201)	
Rate, winsorizing outliers (99th pctl)	2436	0.0314	-0.0124	0.0384	-0.0226	0.088
			(0.0073)		(0.0128)	
Other Serious Violent-Crime Arrests						
Count	2456	0.0544	0.0034	0.0551	0.0063	0.733
			(0.0101)		(0.0177)	
Rate	2436	0.1733	0.1312	0	0.2398	0.236
			(0.1107)		(0.1949)	
Rate, winsorizing outliers (99.5th pctl)	2436	0.1195	0.0321	0.0680	0.0587	0.346
			(0.0341)		(0.0599)	
Rate, winsorizing outliers (99th pctl)	2436	0.0810	0.0116	0.0735	0.0212	0.476
			(0.0163)		(0.0286)	

Notes: For each outcome, the first row replicates our main estimates using counts as a dependent variable. The second row instead uses rates, defined as the number of incidents divided by the number of days not incapacitated, counting any day spent incarcerated or after a homicide victimization as a day when someone is incapacitated. Twenty sample members were incapacitated every day of the 20-month outcome period and are dropped from all rate regressions. The third and fourth rows winsorize rates at the 99.5^{th} and 99^{th} percentiles, respectively, to reduce the influence of outliers created by those individuals with very few non-incapacitated days. Reported means and coefficients for all rate-based regressions are scaled to represent the standard 20-month outcome period (multiplied by 610 days). Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. CM shows control means; ITT shows intent-to-treat estimates; CCM shows control complier means, rounded to zero when estimate is negative due to sampling error; and TOT shows treatment-on-the-treated estimates from an IV regression. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses.

Table A.VII: READI's estimated effects, pre- and post-COVID

	CM		ITT			P-values		
	Pre-COVID	Post-COVID	Treatment Pre-COVID	Treatment Post-COVID	Difference	Treatment Pre-COVID	Treatment Post-COVID	Difference
Primary Outcome Components, Counts								
Shooting & Homicide Victimizations	0.0986	0.1470	0.0010	-0.0331	-0.0341	0.942	0.267	0.303
			(0.0142)	(0.0298)	(0.0331)			
Shooting & Homicide Arrests	0.0268	0.0267	-0.0146	-0.0040	0.0105	0.028	0.746	0.461
			(0.0066)	(0.0124)	(0.0143)			
Other Serious Violent-Crime Arrests	0.0568	0.0468	0.0068	-0.0072	-0.0140	0.563	0.679	0.514
			(0.0118)	(0.0173)	(0.0214)			
All Events, Counts								
All Arrests	1.8808	0.9619	-0.0244	-0.1685	-0.1441	0.804	0.084	0.280
			(0.0984)	(0.0975)	(0.1332)			
All Victimizations	0.5415	0.4776	-0.0191	0.0143	0.0334	0.701	0.834	0.687
			(0.0499)	(0.0681)	(0.0830)			

Notes: N = 2,456. Treatment effects estimated with a person-day panel including post-randomization observations only. Regressions include a post-COVID indicator, an indicator for a treated person-day, and the interaction of the two, in addition to baseline covariates and randomization strata fixed effects. The pre-COVID ITT is the estimated coefficient on the treated person-day indicator. The post-COVID ITT is the linear combination of the estimated coefficients on the treated person-day indicator and its interaction with the post-COVID indicator. The difference is the estimated coefficient on the interaction of the treated person-day and post-COVID indicators. Reported means and coefficients are scaled to represent the standard 20-month outcome period (multiplied by 610 days). The pre-COVID period runs through 3/15/20. The post-COVID period runs starts after 3/15/20. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. CM shows control means; ITT shows intent-to-treat estimates. Heteroskedasticity-robust standard errors clustered by individual reported in parentheses.

Table A.VIII: READI's estimated effects, by time of the week

	Work Hours (Mon-Fri 8am-6pm)			Weekend (Fri 6pm - Sun 11:59pm)			Weekday Mornings and Nights (Mon-Fri 6pm-8am)		
	CM	ITT	P-value	CM	ITT	P-value	CM	ITT	P-value
Primary Outcome Components, Counts									
Shooting & Homicide Victimizations	0.0284	0.0001	0.991	0.0528	-0.0009	0.924	0.0292	-0.0064	0.331
		(0.0068)			(0.0094)			(0.0066)	
Shooting & Homicide Arrests	0.0081	-0.0016	0.640	0.0138	-0.0094	0.015	0.0049	-0.0010	0.733
		(0.0035)			(0.0039)			(0.0028)	
Other Serious Violent-Crime Arrests	0.0154	0.0067	0.233	0.0227	-0.0060	0.338	0.0162	0.0027	0.612
		(0.0056)			(0.0062)			(0.0054)	
All Events, Counts									
All Victimizations	0.1826	-0.0211	0.319	0.2005	0.0201	0.378	0.1429	-0.0101	0.551
		(0.0211)			(0.0228)			(0.0170)	
All Arrests	0.6055	-0.0589	0.164	0.6047	0.0171	0.657	0.4472	-0.0176	0.591
		(0.0423)			(0.0385)			(0.0328)	

Notes: N = 2,456. Estimates from regressions of outcomes reflecting counts only during the days and times shown. Weekday mornings and nights does not include events that occur after 6pm on Friday. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. CM shows control means; ITT shows intent-to-treat estimates. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses.

Table A.IX: READI's estimated effects on primary outcome sub-components and binary outcomes

	CM	ITT	ITT P-value	CCM	TOT
Counts					
Shooting & Homicide Victimizations					
Homicide Victimization	0.0252	-0.0040	0.516	0.0253	-0.0073
		(0.0061)			(0.0108)
Shooting Victimizations	0.0852	-0.0032	0.790	0.0852	-0.0059
		(0.0120)			(0.0212)
Shooting & Homicide Arrests					
Homicide Arrests	0.0227	-0.0129	0.016	0.0342	-0.0237
		(0.0053)			(0.0095)
Shooting Arrests	0.0065	-0.0016	0.602	0.0045	-0.0030
		(0.0031)			(0.0055)
Other Serious Violent-Crime Arrests					
Aggravated Assault & Battery (Non-Shooting) Arrests	0.0325	-0.0047	0.493	0.0415	-0.0086
		(0.0068)			(0.0120)
Robbery Arrests	0.0203	0.0047	0.502	0.0138	0.0086
		(0.0070)			(0.0123)
Sexual Assault Arrests	0.0016	0.0034	0.169	0	0.0063
		(0.0025)			(0.0044)
Indicators					
Shooting & Homicide Victimizations					
Homicide Victimization	0.0252	-0.0040	0.516	0.0253	-0.0073
		(0.0061)			(0.0108)
Shooting Victimization	0.0795	-0.0014	0.895	0.0805	-0.0027
		(0.0110)			(0.0193)
Shooting & Homicide Arrests					
Homicide Arrest	0.0219	-0.0120	0.019	0.0326	-0.0221
		(0.0051)			(0.0091)
Shooting Arrest	0.0065	-0.0016	0.602	0.0045	-0.0030
		(0.0031)			(0.0055)
Other Serious Violent-Crime Arrests					
Aggravated Assault & Battery (Non-Shooting) Arrest	0.0325	-0.0047	0.493	0.0415	-0.0086
		(0.0068)			(0.0120)
Robbery Arrest	0.0170	0.0039	0.469	0.0123	0.0072
		(0.0054)			(0.0095)
Sexual Assault Arrest	0.0016	0.0034	0.169	0	0.0063
		(0.0025)			(0.0044)

Notes: N = 2,456. The top panel shows estimated effects on the sub-categories of arrest counts that comprise each component of the primary index. The bottom panel includes the same outcomes, but measured as binary indicators rather than counts. CM shows control means; ITT shows intent-to-treat estimates; CCM shows control complier means, rounded to zero when estimate is negative due to sampling error; and TOT shows treatment-on-the-treated estimates from an IV regression. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses.

Table A.X: READI's estimated effects on other outcomes

	CM	ITT	ITT P-value	CCM	TOT
Other Arrest & Victimization					
Counts					
Less Serious Violent-Crime Arrests	0.167	0.002	0.903	0.163	0.004
		(0.020)			(0.035)
Property Crime Arrests	0.100	-0.010	0.599	0.101	-0.017
		(0.018)			(0.032)
Drug Crime Arrests	0.376	-0.053	0.140	0.457	-0.098
		(0.036)			(0.064)
Other Crime Arrests	0.939	0.007	0.898	0.959	0.013
		(0.055)			(0.097)
Other Violent Victimizations	0.290	-0.015	0.634	0.323	-0.028
		(0.032)			(0.057)
Indicators					
Less Serious Violent-Crime Arrest	0.125	0.013	0.347	0.121	0.023
		(0.013)			(0.024)
Property Crime Arrest	0.075	-0.001	0.934	0.068	-0.002
		(0.011)			(0.019)
Drug Crime Arrest	0.222	-0.026	0.103	0.261	-0.048
		(0.016)			(0.029)
Other Crime Arrest	0.487	-0.001	0.971	0.507	-0.001
		(0.019)			(0.034)
Other Violent Victimization	0.167	-0.001	0.924	0.182	-0.003
		(0.015)			(0.026)
Incarceration & Incapacitation					
Counts					
Percent of Days Incapacitated	0.144	-0.012	0.170	0.123	-0.023
		(0.009)			(0.016)
Days Incapacitated	87.5	-7.6	0.170	75.3	-13.9
		(5.5)			(9.7)
Days Lost to Homicide	8.2	-3.5	0.078	9.1	-6.4
		(2.0)			(3.5)
Days Incarcerated	79.3	-4.1	0.440	66.2	-7.5
		(5.3)			(9.3)
Days in Jail (CCSO)	40.8	0.022	0.995	33.9	0.040
		(3.5)			(6.2)
Days in Prison (IDOC)	38.5	-4.1	0.248	32.3	-7.5
		(3.6)			(6.3)
Indicators					
Incapacitated	0.485	-0.010	0.590	0.463	-0.018
		(0.018)			(0.033)
Incarcerated	0.471	-0.007	0.712	0.450	-0.012
		(0.018)			(0.033)
Incarcerated in Jail (CCSO)	0.451	-0.017	0.363	0.443	-0.031
		(0.019)			(0.033)
Incarcerated in Prison (IDOC)	0.217	-0.008	0.582	0.202	-0.015
		(0.015)			(0.026)

Notes: N = 2,456. Table shows READI's estimated effects on secondary outcomes (counts and indicators) that measure other involvement in crime and violence. Days Incapacitated measure days during which an individual was either incarcerated or deceased. Days Incarcerated separates incarceration in the Cook County Jail and in a state prison (Illinois Department of Corrections). Less serious violent-crime arrests include non-Part I violent-crime arrests, such as simple assault and battery and domestic violence. Other (non-shooting) violent victimizations include sexual assault, robbery, aggravated assault and battery, and simple assault and battery. CM shows control means; ITT shows intent-to-treat estimates; CCM shows control complier means; and TOT shows treatment-on-the-treated estimates from an IV regression. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses.

Table A.XI: READI's estimated effects on other outcomes, by time of the week

	Work Hours (Mon-Fri 8am-6pm)			Weekend (Fri 6pm - Sun 11:59pm)			Weekday Mornings and Nights (Mon-Fri 6pm-8am)		
	CM	ITT	P-value	CM	ITT	P-value	CM	ITT	P-value
Other Arrest & Victimization Counts									
Less Serious Violent-Crime Arrests	0.0471	-0.0017	0.849	0.0625	0.0210	0.078	0.0576	-0.0168	0.118
		(0.0091)			(0.0119)			(0.0108)	
Property Crime Arrests	0.0317	0.0070	0.501	0.0365	-0.0052	0.527	0.0317	-0.0114	0.159
		(0.0104)			(0.0082)			(0.0081)	
Drug Crime Arrests	0.1656	-0.0455	0.015	0.1274	0.0062	0.745	0.0828	-0.0139	0.286
		(0.0186)			(0.0190)			(0.0130)	
Other Crime Arrests	0.3409	-0.0265	0.369	0.3433	0.0111	0.673	0.2549	0.0225	0.353
		(0.0295)			(0.0262)			(0.0242)	
Other Violent Victimizations	0.1071	-0.0296	0.049	0.1006	0.0192	0.262	0.0820	-0.0049	0.721
		(0.0150)			(0.0171)			(0.0136)	

Notes: N = 2,456. Estimates from regressions of outcomes reflecting counts only during the days and times shown. Weekday mornings and nights does not include events that occur after 6pm on Friday. Less serious violent-crime arrests include non-Part I violent-crime arrests, such as simple assault and battery and domestic violence. Other (non-shooting) violent victimizations include sexual assault, robbery, aggravated assault and battery, and simple assault and battery. CM shows control means; ITT shows intent-to-treat estimates. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses.

Table A.XII: READI's estimated effects on serious violence involvement, by site

	Estimates				P-values		
	CM	ITT	CCM	TOT	Observed ITT	FWER	FDR-q
Primary Index of Serious Violence by Site							
Austin/West Garfield Park (N = 852)	0.0130	-0.0652	0.0478	-0.1141	0.088	0.241	0.164
		(0.0381)		(0.0646)			
Greater Englewood (N = 775)	0.0080	-0.0630	0.0416	-0.1176	0.109	0.241	0.164
		(0.0393)		(0.0702)			
North Lawndale (N = 829)	-0.0209	0.0476	-0.0454	0.0903	0.270	0.268	0.270
		(0.0432)		(0.0790)			
Primary Outcome Components by Site, Counts							
Austin/West Garfield Park							
Shooting & Homicide Victimizations	0.1049	-0.0081	0.1012	-0.0140	0.700	0.731	0.700
		(0.0210)		(0.0354)			
Shooting & Homicide Arrests	0.0350	-0.0208	0.0448	-0.0365	0.046	0.128	0.138
		(0.0104)		(0.0177)			
Other Serious Violent-Crime Arrests	0.0559	-0.0114	0.0699	-0.0201	0.481	0.731	0.700
		(0.0162)		(0.0274)			
Greater Englewood							
Shooting & Homicide Victimizations	0.1114	-0.0263	0.1298	-0.0489	0.254	0.443	0.381
		(0.0230)		(0.0411)			
Shooting & Homicide Arrests	0.0233	-0.0172	0.0368	-0.0321	0.047	0.130	0.142
		(0.0087)		(0.0155)			
Other Serious Violent-Crime Arrests	0.0648	-0.0019	0.0560	-0.0036	0.921	0.921	0.921
		(0.0188)		(0.0335)			
North Lawndale							
Shooting & Homicide Victimizations	0.1151	0.0119	0.1020	0.0224	0.640	0.868	0.831
		(0.0253)		(0.0464)			
Shooting & Homicide Arrests	0.0216	0.0024	0.0187	0.0044	0.831	0.868	0.831
		(0.0111)		(0.0203)			
Other Serious Violent-Crime Arrests	0.0432	0.0231	0.0390	0.0440	0.191	0.472	0.573
		(0.0177)		(0.0323)			

Notes: N = 2,456. Estimates for each outcome are from a single regression that interacts site indicators with treatment. Primary index standardizes each of the three components shown in the bottom panel based on the control group's distribution and takes their unweighted average. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. For multiple hypothesis testing adjustments within the primary index, we define the three sites as a family. For the component adjustments, we define the three different outcomes within each site as a family. FWER p-values control for the family-wise error rate using the Westfall and Young (1993) method for step-down resampling. FDR-q values control for the false discovery rate using the Benjamini and Hochberg (1995) method. CM shows control means; ITT shows intent-to-treat estimates; CCM shows control complier means; and TOT shows treatment-on-the-treated estimates from an IV regression. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses. F-tests of the null hypothesis that all three treatment effects are equal across sites are as follows: Primary Index: $p = 0.090$; Shooting & Homicide Victimizations: $p = 0.535$; Shootings & Homicide Arrests: $p = 0.250$; Other Serious Violent-Crime Arrests: $p = 0.342$

Table A.XIII: READI's estimated effects on serious violence involvement, by age group

	Estimates				P-values		
	CM	ITT	CCM	TOT	Observed ITT	FWER	FDR-q
Primary Index of Serious Violence by Age							
Over Median (N = 1228)	-0.0628	-0.0099	-0.0734	-0.0178	0.726	0.721	0.726
		(0.0284)		(0.0502)			
Under Median (N = 1228)	0.0637	-0.0399	0.1001	-0.0744	0.293	0.500	0.585
		(0.0380)		(0.0682)			
Primary Outcome Components by Age, Counts							
Over Median							
Shooting & Homicide Victimizations	0.0887	-0.0039	0.0872	-0.0071	0.825	0.940	0.825
		(0.0176)		(0.0312)			
Shooting & Homicide Arrests	0.0145	0.0023	0.0043	0.0046	0.753	0.940	0.825
		(0.0073)		(0.0130)			
Other Serious Violent-Crime Arrests	0.0419	-0.0079	0.0504	-0.0148	0.497	0.870	0.825
		(0.0117)		(0.0207)			
Under Median							
Shooting & Homicide Victimizations	0.1324	-0.0093	0.1321	-0.0173	0.648	0.644	0.648
		(0.0204)		(0.0367)			
Shooting & Homicide Arrests	0.0392	-0.0252	0.0621	-0.0470	0.010	0.029	0.030
		(0.0098)		(0.0176)			
Other Serious Violent-Crime Arrests	0.0670	0.0146	0.0602	0.0274	0.383	0.624	0.574
		(0.0167)		(0.0301)			

Notes: N = 2,456. Estimates for each outcome are from a single regression that interacts age group with treatment. Median age at time of randomization is 24.7. Primary index standardizes each of the three components shown in the bottom panel based on the control group's distribution and takes their unweighted average. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. For multiple hypothesis testing adjustments within the primary index, we define the two age groups as a family. For the component adjustments, we define the three different outcomes within each age group as a family. FWER p-values control for the family-wise error rate using the Westfall and Young (1993) method for step-down resampling. FDR-q values control for the false discovery rate using the Benjamini and Hochberg (1995) method. CM shows control means; ITT shows intent-to-treat estimates; CCM shows control complier means; and TOT shows treatment-on-the-treated estimates from an IV regression. Regressions include baseline covariates and randomization strata fixed effects. Heteroskedasticity-robust standard errors reported in parentheses. F-tests of the null hypothesis that all both treatment effects are equal across age groups are as follows: Primary Index: $p = 0.529$; Shooting & Homicide Victimizations: $p = 0.841$; Shootings & Homicide Arrests: $p = 0.027$; Other Serious Violent-Crime Arrests: $p = 0.275$

Table A.XIV: Predicting outreach referrals

	Outreach Referral
Demographics	
Non-Black (Other Race/Ethnicity or Missing)	-0.124***
Age (std.)	-0.004
Primary Outcome Components	
Shooting Victimizations (std.)	-0.007
Shooting & Homicide Arrests (std.)	-0.005
Other Serious Violent-Crime Arrests (std.)	-0.006
Risk Prediction	
Risk Score (Zero-Imputed) (std.)	-0.152***
Missing Risk Score	0.107***
Arrest Counts	
Less Serious Violent-Crime Arrests (std.)	-0.041***
Drug Crime Arrests (std.)	-0.013
Property Crime Arrests (std.)	-0.003
Other Crime Arrests (std.)	-0.018
Victimization Counts	
Other Violent Victimizations (std.)	-0.019
Non-Violent Victimizations (std.)	-0.014
Incarceration Measures	
Days Incarcerated (std.)	-0.059***
Incarcerated at Randomization	0.021
Site	
North Lawndale	0.028
Greater Englewood	-0.035
Adjusted R-squared	0.168

Notes: N = 2,456. Table reports coefficients from a regression of an indicator for being in the outreach pathway on baseline covariates and site indicators. Covariates with "(std.)" have been standardized by the variable's standard deviation, so coefficient magnitudes correspond to a one standard deviation increase in the covariate. All other covariates are indicator variables. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. Less serious violent-crime arrests include non-Part I violent-crime arrests, such as simple assault and battery and domestic violence. Other (non-shooting) violent victimizations include sexual assault, robbery, aggravated assault and battery, and simple assault and battery. Non-violent victimizations include all other incidents such as burglary, stalking, and threats. All regressions estimate heteroskedasticity-robust standard errors. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.XV: Outreach pathway and re-weighted ITT estimates

		Algorithm: Common Support			Re-entry: Common Support		
	Outreach	Outreach	Algorithm		Outreach	Re-entry	
	ITT	ITT	ITT	Implied ITT with Similar Xs	ITT	ITT	Implied ITT with Similar Xs
N	878	462	1186	1186	786	338	338
Primary Index of Serious Violence							
	-0.1007	-0.2158	0.0422	0.0847	-0.1087	-0.0721	0.0907
Primary Outcome Components, Counts							
Shooting & Homicide Victimizations	-0.0438	-0.0738	0.0266	0.0131	-0.0444	-0.0300	0.0134
Shooting & Homicide Arrests	-0.0255	-0.0452	-0.0054	0.0099	-0.0268	-0.0042	0.0238
Other Serious Violent-Crime Arrests	-0.0045	-0.0387	0.0197	0.0331	-0.0081	-0.0252	0.0061

Notes: For the entire sample, we first estimate predicted probabilities of being an outreach referral, $\hat{p}_i^{outreach}$, using a logit on our standard set of baseline covariates (without strata fixed effects). Then, we use visual inspection of the densities of $\hat{p}_i^{outreach}$ by pathway to determine where common support holds. Enforcing common support between outreach and algorithm referrals results in dropping 416 outreach and 46 algorithm referrals. Enforcing common support between outreach and re-entry referrals results in dropping 92 outreach and 8 re-entry referrals. Finally, the "Implied ITT with Similar Xs" columns implement the correction for observables method described in Andrews and Oster (2021), estimating what the intent-to-treat effect would be among algorithm and re-entry referrals, respectively, if they had the mean observables of outreach referrals. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. For more details, see Appendix A.6.3.

Table A.XVI: Predicting take-up overall and within pathway

	Take-up Indicator	Take-up Indicator	Hours Worked	Hours Worked
N	1224	1224	551	551
Demographics				
Non-Black (Other Race/Ethnicity or Missing)	-0.097	-0.042	-282.3***	-279.0***
Age (std.)	-0.001	-0.002	39.7	41.1
Primary Outcome Components				
Shooting Victimizations (std.)	-0.028*	-0.028*	19.1	19.9
Shooting & Homicide Arrests (std.)	-0.013	-0.008	-12.8	-12.8
Other Serious Violent-Crime Arrests (std.)	0.014	0.006	-20.1	-20.3
Risk Prediction				
Risk Score (Zero-Imputed) (std.)	-0.079***	-0.008	-71.2**	-59.1*
Missing Risk Score	-0.028	-0.077	-32.3	-34.4
Arrest Counts				
Less Serious Violent-Crime Arrests (std.)	-0.037**	-0.017	-11.7	-9.8
Drug Crime Arrests (std.)	0.028	0.037**	9.6	13.7
Property Crime Arrests (std.)	-0.025*	-0.023*	-42.0**	-43.5**
Other Crime Arrests (std.)	-0.020	-0.010	5.4	6.5
Victimization Counts				
Other Violent Victimizations (std.)	-0.002	0.009	21.7	23.7
Non-Violent Victimizations (std.)	-0.008	-0.006	-47.5**	-47.8**
Incarceration Measures				
Days Incarcerated (std.)	0.009	0.010	-89.3***	-90.1***
Incarcerated at Randomization	-0.147**	-0.128*	-29.0	-32.8
Site				
North Lawndale	-0.061*	-0.056*	9.8	12.3
Greater Englewood	-0.029	-0.015	38.7	41.6
Pathway				
Outreach Pathway		0.406***		65.2
Re-entry Pathway		0.200***		48.9
Adjusted R-squared	0.046	0.153	0.065	0.065
D.V. Mean	0.548	0.548	541.8	541.8

Notes: First two columns regress an indicator for ever attending READI orientation on the covariates listed, using just the treatment group. Second two columns regress the number of hours worked in the payroll data on the covariates listed, using only participants with non-missing payroll data. "D.V. Mean" reports the mean of each dependent variable. Covariates with "(std.)" have been standardized by the variable's standard deviation, so coefficient magnitudes correspond to a one standard deviation increase in the covariate. All other covariates are indicator variables. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. Less serious violent-crime arrests include non-Part I violent-crime arrests, such as simple assault and battery and domestic violence. Other (non-shooting) violent victimizations include sexual assault, robbery, aggravated assault and battery, and simple assault and battery. Non-violent victimizations include all other incidents such as burglary, stalking, and threats. All regressions estimate heteroskedasticity-robust standard errors. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.XVII: Heterogeneity by predicted dosage

	Low predicted dosage			High predicted dosage		
	CM	LOO	RSS	CM	LOO	RSS
Primary Index of Serious Violence						
	0.0103	0.0009	-0.0100	-0.0107	-0.0551	-0.0395
		(0.0399)	(0.0327)		(0.0347)	(0.0281)
Primary Outcome Components, Counts						
Shooting & Homicide Victimizations	0.1048	0.0064	0.0007	0.1163	-0.0277	-0.0148
		(0.0225)	(0.0194)		(0.0203)	(0.0171)
Shooting & Homicide Arrests	0.0286	-0.0047	-0.0042	0.0249	-0.0112	-0.0176**
		(0.0100)	(0.0082)		(0.0091)	(0.0075)
Other Serious Violent-Crime Arrests	0.0635	0.0029	-0.0017	0.0449	-0.0038	0.0077
		(0.0180)	(0.0144)		(0.0155)	(0.0127)
Take-up rate		39%			70%	
Average total hours		143			426	
N		1,228			1,228	

Notes: Table shows leave-one-out (LOO) and repeated split sample (RSS) ITT estimates from a method adapted from Abadie et al. (2018). The analysis predicts total hours of participation among the treatment group, then uses the prediction regression to assign a predicted dosage for the out-of-sample observation(s), both treatment and control. Low and high predicted dosage groups are split at the median prediction. Other serious violent-crime arrests include criminal sexual assault, robbery, and non-shooting aggravated assault and battery. CM shows control means. Regressions include baseline covariates and randomization strata fixed effects. Standard errors from methods in Abadie et al. (2018). For more details, see Appendix A.6.5. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.XVIII: Inputs to social cost of crime estimates

Crime Type	Bottom-Up Components: Victim Costs	Bottom-Up Components: Legal System Costs	Bottom-Up Components: Offender Productivity Costs	Total Bottom-Up Costs	Willingness-to-Pay Costs	Estimated Clearance Rate	Estimated Reporting Rate
Murder	5,438,106	354,659	165,508	5,910,985	13,429,758	0.37	1.00
Rape	159,597	9,812	5,320	177,330	327,705	0.28	0.32
Armed Robbery	34,284	17,378	9,458	59,110	304,179	0.09	0.52
Robbery	14,186	8,748	4,729	27,191	32,629	0.09	0.52
Aggravated Assault	43,741	15,960	7,566	65,021	76,961	0.15	0.56
Simple Assault	5,320	5,911	1,537	13,004	15,014	0.11	0.37
Burglary	2,364	2,719	1,182	5,911	37,476	0.09	0.48
Motor Vehicle Theft	6,502	3,428	1,182	10,640	15,487	0.12	0.78
Larceny	532	2,010	828	3,310	1,892	0.11	0.28
Drunk Driving Crash	33,102	2,010	828	35,466	68,095	0.09	1.00
Arson	67,385	2,010	828	70,932	133,115	0.09	1.00
Vandalism	437	745	0	1,182	1,620	0.09	0.28
Fraud	1,300	2,010	828	4,138	3,665	0.09	0.28
Other	0	591	0	591	591	0.09	0.28

Notes: Crime types, bottom-up components, and willingness-to-pay costs are based on estimates from Cohen and Piquero (2009). We inflate to 2017 dollars and subtract legal system and offender productivity costs from willingness-to-pay estimates to avoid double counting. Total bottom-up costs are the result of summing up the bottom-up component costs. Estimated clearance rate for murder from CPD. Estimated clearance rates for other offenses are the midpoints of the ranges reported in Belfield et al. (2006). Rates of reporting victimization to the police are derived from Morgan and Thompson (2021).

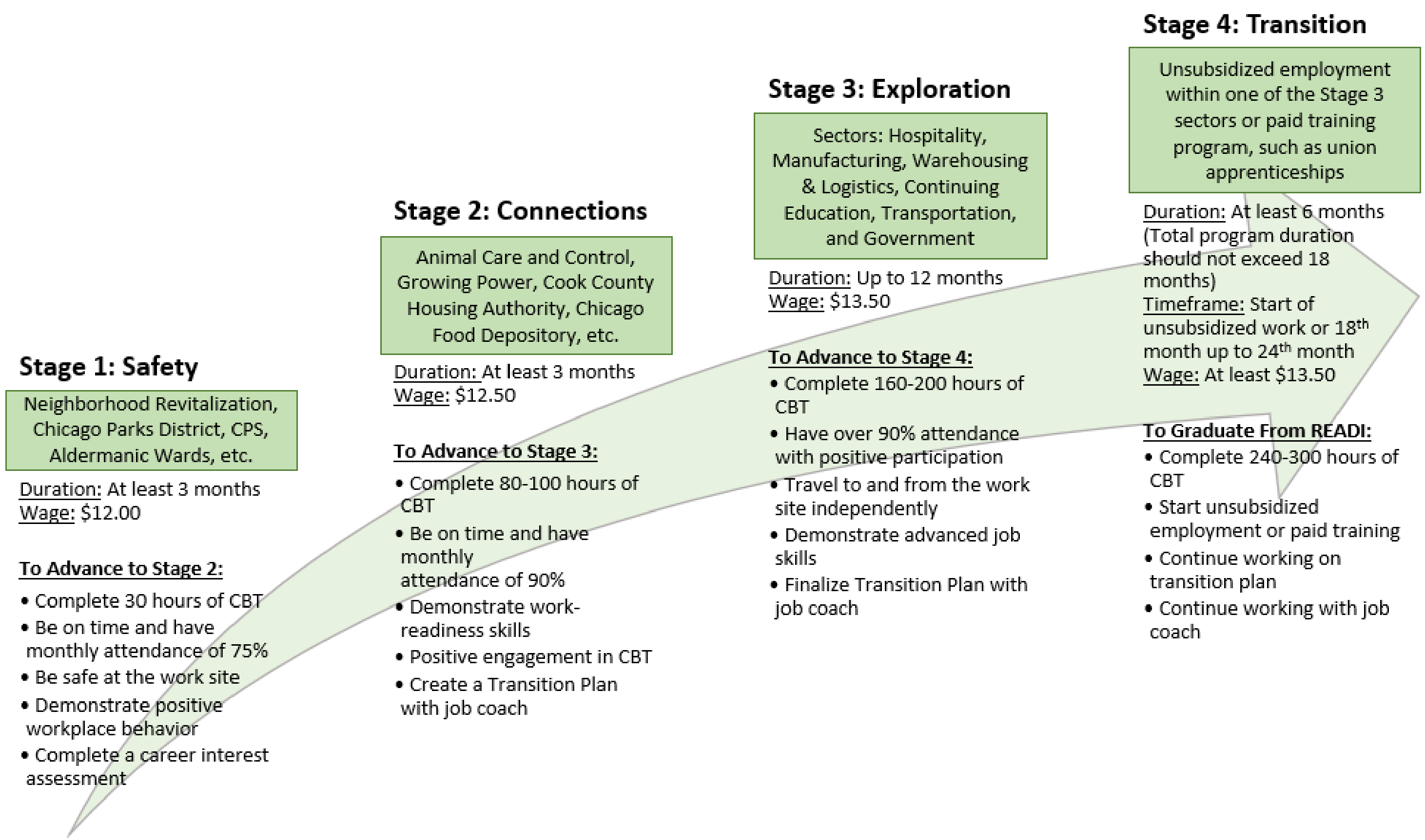
Table A.XIX: Percentage of focus group respondents identifying each READI component as the most important

	All	Outreach Workers	CBT Facilitators	Work Crew Chiefs	Managers
Relentless Engagement (Outreach)	15.5	25.8	17.4	0	6.7
Work	10.7	19.4	0	20.0	0
CBT	19.1	22.6	21.7	20.0	6.7
Some Combination	34.5	29	34.8	53.3	26.7
Responded to Question	49/84	19/31	16/23	9/15	5/15

Notes: Table reports the proportion of staff focus group respondents, overall and by role, who voted for each READI component as being the most important. Not all focus groups were asked this question, and some focus group participants chose not to respond.

Figure A.I

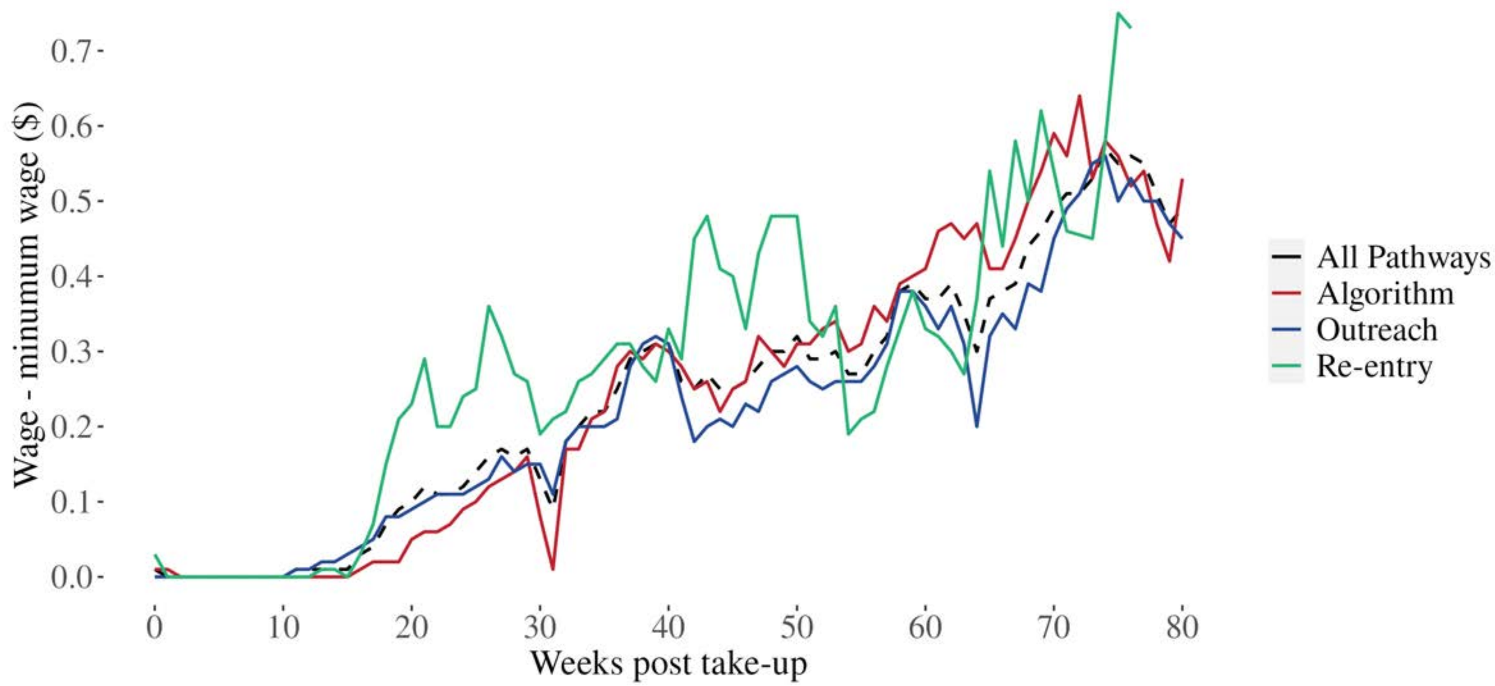
READI job stage progression



Notes: The duration and advancement requirements for READI's job stages were subject to change based on a participant's needs and progress. Program staff exercised discretion in deciding which participants were ready to advance job stages. The diagram shows READI's initial design; the details of implementation varied somewhat in practice as the model developed over time.

Figure A.II

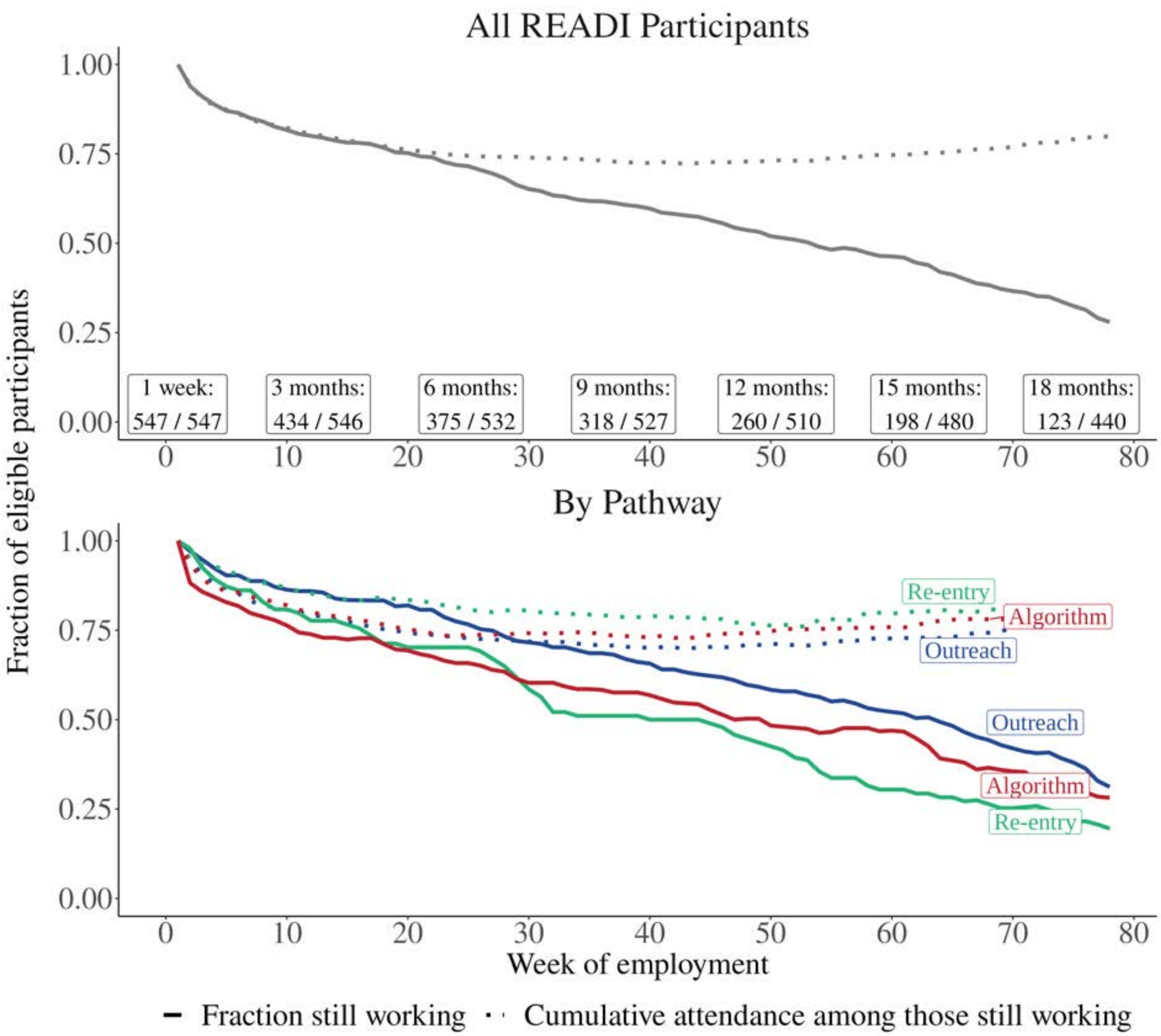
READI wage growth by pathway



Notes: READI's starting wage was $11 at its launch in August 2017, increased to $12 in July 2018, and to $13 in July 2019. Average wage is calculated using only participants who report to work during a given week.

Figure A.III

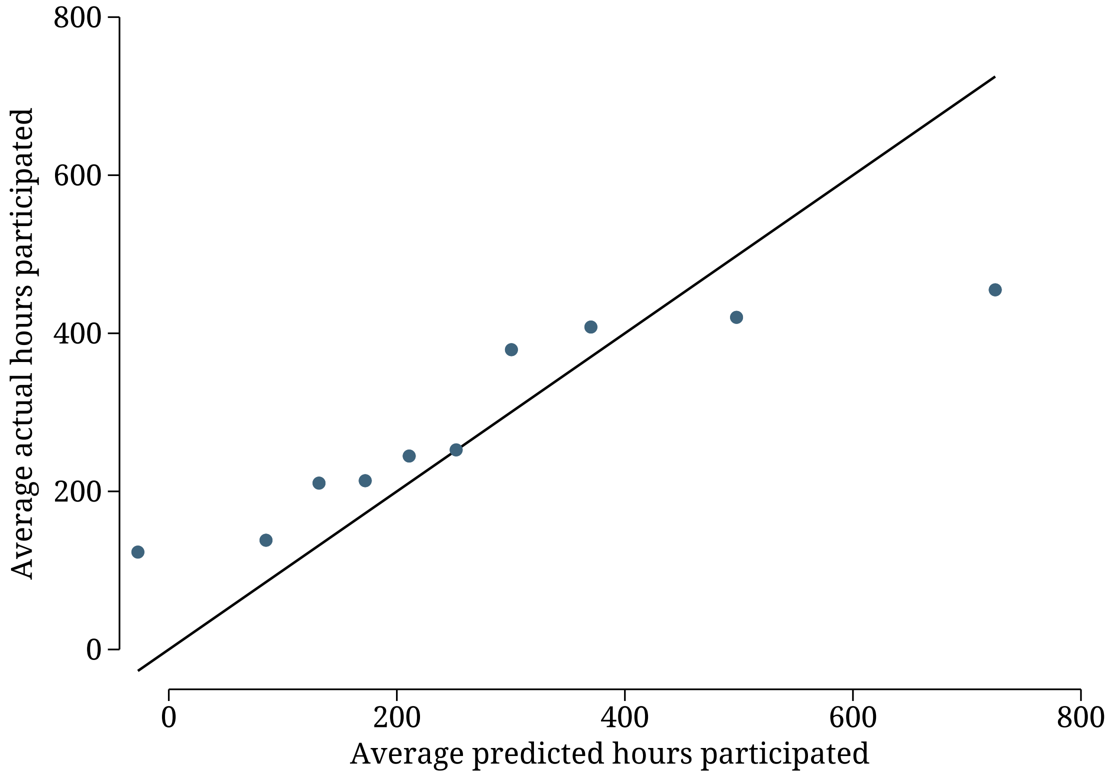
READI job retention, overall and by pathway, including COVID period



Notes: Figure shows two measures of job retention for men who started READI measured from payroll data. The solid line shows the proportion of participants who work at least once after the time shown on the x-axis conditional on observing them for that long. The boxes show the number of workers contributing to each point. The dotted line shows the average proportion of possible weeks worked among those still working at each point in time. At 18 months after first taking up, N = 38 algorithm referrals, N = 68 outreach referrals, and N = 17 re-entry referrals are still observed working.

Figure A.IV

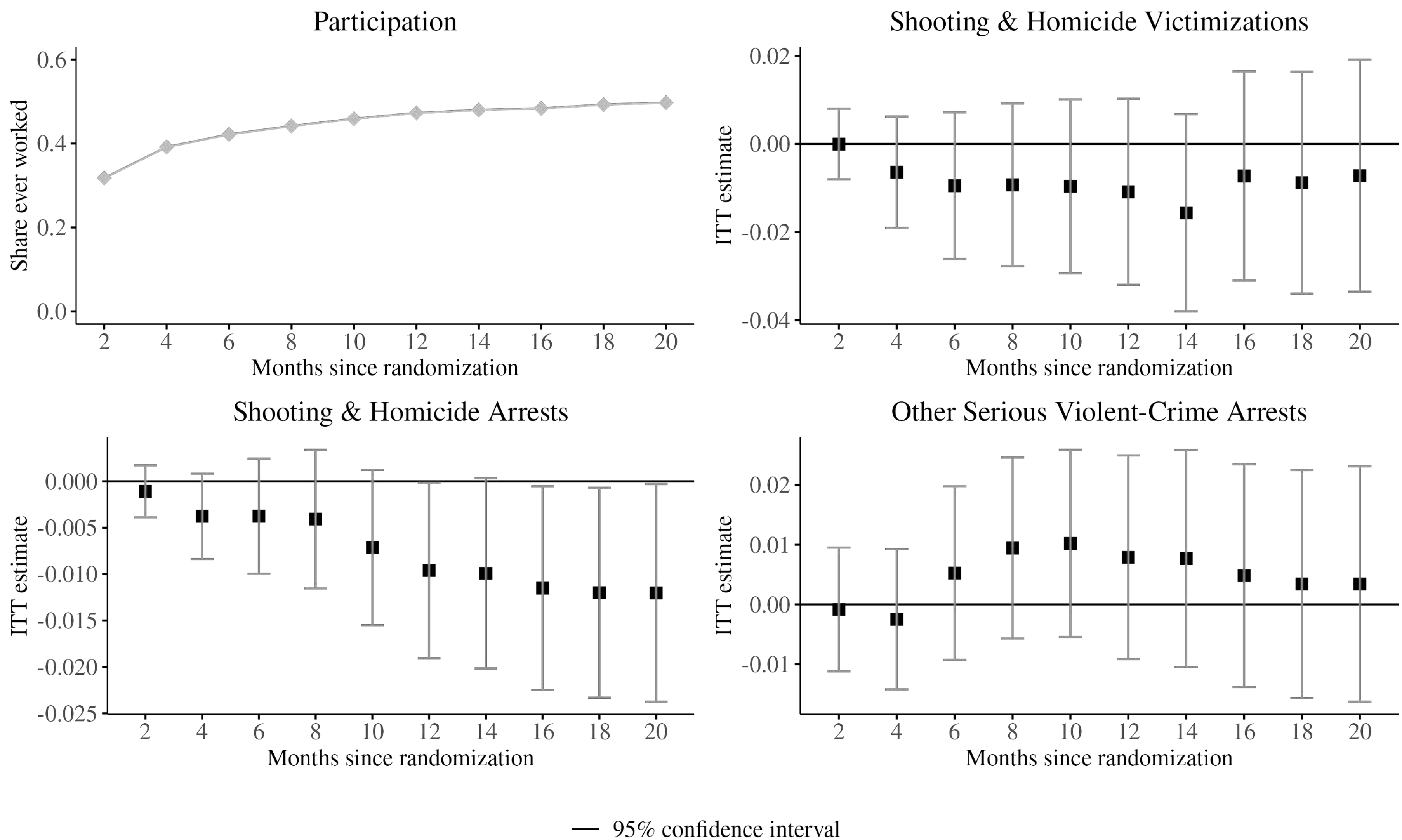
Predicted and actual dosage for treatment participants



Notes: Figure shows dosage predictions from an endogenous stratification exercise that uses a leave-one-out regression of hours participated on observable baseline covariates among the treatment group. X-axis is the average predicted hours participated in each decile bin. Y-axis is the average true hours participated in that bin. The solid line is the 45 degree line that would reflect good calibration across the distribution. Appendix Section A.6.5 discusses the use of these predictions to analyze treatment heterogeneity by predicted dosage for the entire sample.

Figure A.V

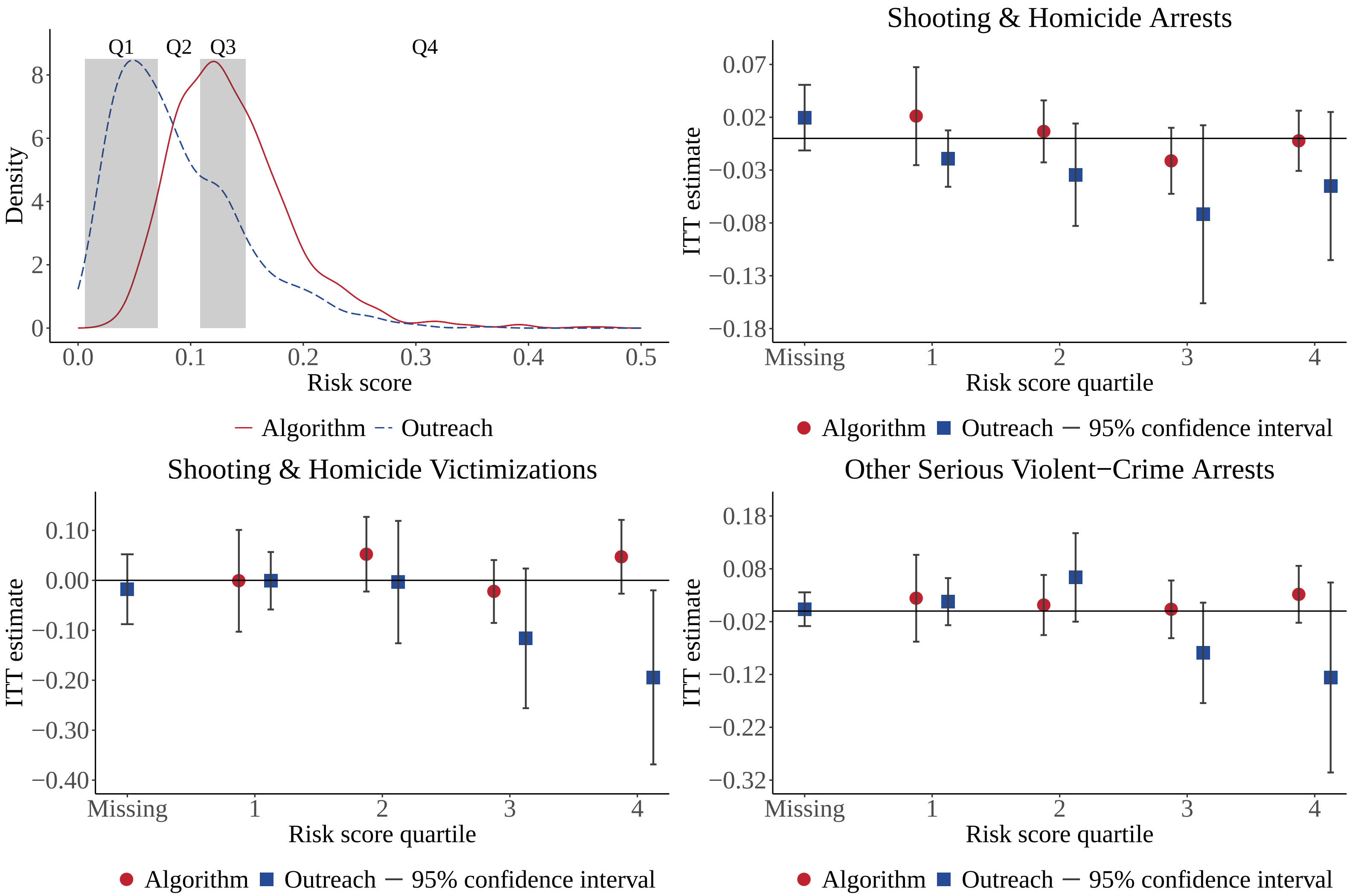
Cumulative first stage and ITT effects over time



Notes: Figures show cumulative treatment effects up to the time shown on the x-axis, inclusive. Top left panel shows indicators for any participation; other panels show the three main components of the primary index. Regressions include baseline covariates and randomization strata fixed effects, and 95 percent confidence intervals are constructed using heteroskedasticity-robust standard errors.

Figure A.VI

Distribution of risk scores by pathway and estimated effects on index components by pathway and risk level



Notes: Top left panel shows distributions of the risk score, the predicted probability at baseline of being a victim or an arrestee in a violent gun crime during the next 18 months, by pathway. Shaded areas denote quartiles of the risk score. The remaining panels show coefficient estimates and 95 percent confidence intervals (using heteroskedasticity-robust standard errors) on three-way interactions of pathway indicators, risk quartile indicators, and an indicator for being randomized to receive a READI offer, from regressions of the primary index components on baseline covariates, randomization strata fixed effects, and all two-way interactions of pathway indicators, risk quartile indicators, and an indicator for being randomized to receive a READI offer.