

A Community-Response Approach to Mental-Health and Substance-Abuse Crises Reduced Crime

Thomas S. Dee*

Jaymes Pyne*

Stanford University

October 18, 2021

Abstract

Police officers often serve as first responders to mental-health and substance-abuse crises. Concerns over the unintended consequences and high costs associated with this approach have motivated new emergency-response models that augment or completely remove police involvement. However, there is little causal evidence evaluating these programs. This pre-registered study presents quasi-experimental evidence on the impact of an innovative “community response” pilot in Denver that directed targeted emergency calls to health-care responders instead of the police. Tracking reported criminal incidents regardless of who responded or whether they led to formal charges, we find robust evidence that the program reduced targeted, less serious crimes (e.g., trespassing, public disorder, resisting arrest) by 34 percent and had no detectable effect on more serious crimes. We also find that this community-response approach has substantially lower costs relative to engaging such crises through the criminal-justice system.

*Corresponding authors. Emails: tdee@stanford.edu; pyne@stanford.edu. Authors contributed equally to this manuscript.

Police often serve as first responders to emergency calls involving non-violent individuals in mental-health distress or suffering from alcohol or drug abuse. This procedural norm has been the subject of debate and criticism for two broad reasons. One is that serving as first responders to calls involving mental-health crises is a substantial drain on scarce police resources. Police currently spend more time responding to such “low priority” calls than to any other type of emergency call (1). Recent estimates (2) suggest that one to two-thirds of the emergency calls to which police currently respond could instead be directed to mental-health crisis experts (i.e., a “community response” model). Those charged with minor offenses like loitering, making false statements and vandalism cost the criminal-justice system roughly \$500-\$600 per offense—and come with even higher additional social costs (3). The potential reallocation of resources away from a police response and towards mental-health supports is often a part of current initiatives to “defund the police” (4).

Second, having armed and uniformed police as first responders to a mental-health or substance-abuse crisis may increase the likelihood of costly outcomes and inappropriate care. Individuals living with serious mental illness are no more prone to violence or unpredictability than the general population (5, 6). However, having police officers as first responders to a mental-health crisis can result in unnecessarily violent and tragic outcomes (7, 8). Recent news coverage has drawn public attention to particularly shocking incidents in which responding police officers seriously harmed or killed a person in mental-health distress (9-11). More generally, having police respond to such incidents can be costly and unproductive because police are more likely than mental-health clinicians to direct individuals experiencing a mental-health episode to the criminal-justice system rather than to the appropriate health care (7).

In response to these concerns, municipalities across the country have begun to pilot targeted reforms. The two most common approaches augment the capacity of police officers to serve as effective first responders to individuals experiencing mental-health crises. The “Crisis Intervention Team” (CIT) approach emphasizes training

police officers how to respond to individuals in crisis and connect them with appropriate services (12). In contrast, the “Co-Response” model involves structuring explicit partnerships between police departments and professional mental health practitioners so they can simultaneously respond to incidents involving mental-health crises (13). A third and less common approach either delays or foregoes on-scene police involvement in certain incidents. These “Community Response” programs establish a triage protocol under which emergency calls for mental-health crises are first addressed by a health team (e.g., a mental-health crisis interventionist and a paramedic) prior to deciding whether to request direct police involvement (2).

The momentum behind the adoption of programs that seek to improve police interactions with individuals in mental-health crises has motivated multiple empirical studies that seek to understand their impact. Several systematic reviews and meta-analyses have synthesized this evidence, particularly focusing on the more common Crisis Intervention Team and Co-Response models (13-18). In general, this empirical literature suggests that these program innovations have positive effects on outcomes such as arrests and detention rates, but evidence is mixed on whether these programs are cost effective. However, the research designs employed in these studies (e.g., case notes, qualitative and descriptive studies, pre/post comparisons, and cross-sectional comparisons) generally do not support credible causal inference. For example, one recent prominent review concludes that “... we caution against drawing conclusions related to causality based on these findings” (17) (p. 12). There is a similar lack of evidence on the impact of less-common Community Response models. Existing evaluations are typically conducted internally by cities, police departments or Community Response teams and rely on descriptive evidence of the number of calls taken by the few Community Response units operating across the U.S. (2). Furthermore, critics warn that initiatives to reduce police involvement in response to emergency calls will “embolden the bad guys” (19) and unintentionally increase the prevalence of more serious criminal offenses (19-21). Such characterizations of the

existing evidence underscore the need for research studies that can provide credibly causal estimates of the impact of these innovative programs both on the focal, less serious crimes they target and on more serious offenses. This study seeks to provide such evidence by examining the impact of a Community Response program recently piloted in the City and County of Denver, Colorado through the independent analysis of a pre-registered, quasi-experimental design coupled with several complementary robustness checks.

The Support Team Assistance Response Pilot (STAR) Program

The Support Team Assistance Response (STAR) program in Denver provides a mobile crisis response for community members experiencing problems related to mental health, depression, poverty, homelessness, and/or substance abuse issues. The STAR response consists of two health-care staff (i.e., a mental-health clinician and a paramedic in a specially equipped van) who provide rapid, on-site support to individuals in crisis and direct them to further appropriate care including requesting police involvement, if necessary. The design of the STAR program is based on the Crisis Assistance Helping Out On The Streets (CAHOOTS) program developed in Eugene, Oregon (22).

STAR began operations on June 1, 2020 for a designated six-month pilot period. During this period, STAR limited its operations to selected 911 calls for assistance in eight purposefully chosen police precincts (i.e., out of the city's 36 precincts), where the need for STAR services were anticipated to be the greatest. The pilot area was in the central downtown area of Denver (Figure S1) and largely represents neighborhoods with residents who are more affluent, educated and white than the city as a whole (see Table S1). However, all but one of the neighborhoods in the STAR pilot service area are also designated by the city as “displacement-vulnerable” areas, a rapidly gentrifying city space where poor and otherwise at-risk residents are being pushed out (23). In such contested urban spaces, there are often

increasing demands on police to conduct “rabble management” that addresses overwhelmingly non-violent incidents (24-27).

Operators responding to 911 calls for assistance dispatched STAR staff to eligible incidents that were located in the designated police precincts and during the program’s hours of operation (Monday through Friday, 10AM to 6PM). The identification of emergency calls eligible for STAR services relied on two specific screening criteria. First, the incident had to designate at least one of several codes: calls for assistance, intoxication, suicidal series, welfare checks, indecent exposure, trespass of an unwanted person, and syringe disposal (28). Second, to dispatch the STAR van, there needed to be no evidence that the incident involved criminal activity, disturbance, weapons, threats, violence, injuries, or serious medical needs. The STAR team also responded to calls from uniformed police to engage with community members in crisis and initiated engagement in the field on their own. Over the six-month pilot period, the STAR team responded to 748 incidents, or nearly six incidents per eight-hour shift (29).

Measuring STAR Impacts on Crime

We identify the impact of the STAR program on STAR-related and unrelated crimes using “difference in differences” (DD) and “difference in difference in differences” (DDD) designs that effectively rely on pre-post comparisons across treated and comparison precincts (i.e., along with the evidence from several complementary robustness checks and alternative estimation procedures). To identify the impact of the STAR program, we consider all criminal offenses reported by the City and County of Denver through data collected as part of their participation in the federal National Incident-Based Reporting System (NIBRS). These data include offenses reported by the police regardless of whether they led to formal charges (including arrest) or whether the STAR team was dispatched or responded to the call. Prior to our analysis, we coded each offense as directly related to STAR operations

(e.g., disorderly conduct, trespassing, alcohol and drug use) or not (e.g., burglary; see Table S2, Table S3, and SI text for details).

The impact of the STAR program on the frequency of these offenses is theoretically uncertain. For example, to the extent police who respond to mental-health incidents consistently direct offending citizens to mental-health care rather than recording criminal offenses, the overall effects of the STAR program would be muted—or even null. In fact, the Denver Police have participated in “Crisis Intervention Team” (CIT) training designed to support their capacity to identify individuals who need mental-health supports and to direct those individuals to appropriate care. Because police officers received CIT training, the introduction of the STAR team should, in theory, have no effect on recorded crimes.

STAR’s “Community Response” (CR) approach could thus reduce crimes through two broad mechanisms. First, having mental-health specialists as first responders to an incident that has already occurred (e.g., trespassing) could result in an increased likelihood of directing offending individuals to mental-health care rather than recording an offense and directing them to the criminal justice system. The STAR team may be more effective than CIT-trained police in identifying mental-health crises and in directing such individuals to appropriate health-care instead. Relatedly, the presence of the STAR program in a precinct could also increase the likelihood that CIT-trained police direct individuals in mental-health crises to health care rather than the criminal-justice system. Second, the STAR program may also reduce the number of criminal incidents in the near future by providing mental-health care to individuals in crisis who might otherwise offend repeatedly. Because individuals in mental-health distress are likely to offend repeatedly, the increased provision of health care could reduce the prevalence of future incidents that would otherwise be recorded as crimes.

We offer indirect evidence on the empirical relevance of these mechanisms by examining the effect of the STAR program on crimes occurring outside their operating hours and by comparing the overall crime-reduction impact of STAR to the number

of incidents to which they directly responded. Similarly, when mental-health clinicians are first responders, they may sometimes find individuals who are committing clear and potentially violent offenses and make the decision to call in the police, resulting in an unchanged reporting of offenses. In contrast, if the STAR program reduces the number of criminal offenses, it implies that mental-health first responders are more effective than police at managing these targeted emergency calls in ways that privilege the provision of health care and that do not engage the criminal-justice system.

Results

As an initial and unrestrictive way to visualize the impact of the STAR program, Figure 1 shows precinct-level maps that illustrate the pre-post changes in crimes by offense type. These maps show that treated precincts experienced sharp, comparative declines in STAR-related crimes but not in those crimes not directly related to STAR services. Estimates based on the DD and DDD designs (i.e., equations (1) and (2)) allow us to estimate these effects directly and to condition on fixed effects unique to precincts and months. Figure 2 displays the key estimates (and corresponding 95% confidence intervals) based on these quasi-experimental specifications.

For our main confirmatory hypothesis, the DD estimate indicates that the STAR program led to large and statistically significant reduction in the targeted crimes ($b = -0.41$, $SE = 0.07$, $t = -6.09$, $p < 0.001$). This estimated impact on the natural log of STAR-related crimes implies that the program reduced these targeted crimes by 34 percent (i.e., $e^{(-0.41)} - 1$). By contrast, the estimated effect of the STAR program on crimes that were *not* directly related to STAR services was comparatively small and statistically insignificant ($b = -0.05$, $SE = 0.04$, $t = -1.18$, $p = 0.245$). This finding suggests that that the targeted fielding of mental-health professionals as first-responders did not increase the frequency of more serious criminal incidents in treated precincts. This null result can also be understood as affirming the causal warrant of the DD design by indicating there were not unobserved and confounding

determinants of crime unique to the precincts and months associated with the STAR pilot, a finding consistent with the causal warrant of the DD design. The DDD specification (i.e., equation (2)) leverages these comparative results by using the data on crimes unrelated to STAR operations as a comparison condition unique to each precinct-month observation. The DDD estimate similarly implies that STAR operations led to a large and statistically significant reduction in the targeted crimes ($b = -0.36$, $SE = 0.05$, $t = -6.64$, $p < 0.001$). This DDD estimate suggests that the STAR program led to a 30 percent reduction [i.e., $e^{(-0.36)} - 1$] in STAR-related offenses (see Table S4 for full numeric results).

Figure 3 illustrates the key estimates from event-study DD specifications that allow for effects unique to treated precincts in each month before and after the onset of STAR operations (see Supplementary Materials Appendix for specification details). The lower red line in Figure 3 represents the point estimates from an unrestrictive event-study specification that examines the treatment-comparison difference in STAR-targeted crimes in the months before and after program implementation (see Table S5 for numeric results). The higher yellow line presents similar estimates based on STAR-unrelated offenses. These results indicate that, in the months before STAR operations, the treated and comparison precincts had similar trends in STAR-related crimes. More formally, we do not reject the hypothesis that the effects on STAR-related crimes that are unique to treated precincts in the months before STAR operations are the same as those in the comparison precincts ($p = 0.71$). These results are consistent with the “parallel trends” assumption of the DD design and with a causal interpretation of the results based on equation (1). The event-study estimates in Figure 3 also illustrate the distinct drop in STAR-related crimes associated with the onset of STAR services as well as the comparative absence of any relationship with the prevalence of crimes that are not directly related to STAR operations. Figure S2 similarly presents conditional means for each group of offenses (i.e., STAR-related and

unrelated) by month for both treatment and comparison precincts, and Figure S3 shows trends in STAR-related offenses among the eight treated districts.

The Supplementary Materials (SM) Appendix presents several ancillary analyses that explore the robustness of these findings. For example, Poisson and negative-binomial specifications that recognize both the count nature of the crime data and the presence of fixed effects (30) return results similar to those based on ordinary least-squares estimates of equation [1] (Table S6). The DD results presented here are also similar in specifications that rely only on when treating May 2020 as a treatment month among STAR-active precincts, in order to allow for anticipation effects (Table S6). When we remove offenses that are STAR-related but not STAR-targeted (i.e., simple assault, simple assault on a police officer, and disarming a police officer; see SI text for details), the static DD effect size is larger than what we report. That is, the point estimate is -0.41 when those assault offenses are included as STAR-related offenses and is -0.49 when they are not (Table S6; see SI text for additional details). We also find similar levels of statistical significance when we remove police precinct 311, which is not entirely serviced by the program, and in specifications that correct for the potential finite-sample bias in the precinct-specific clustering of the error term (Table S6). We also find the results are robust when using permutation-based randomization inference (Figure S4).

Next, to test whether common seasonal changes in crime rates threaten the causal interpretation of these results, we construct parallel “placebo effect” datasets of months from December 2016 - November 2017, December 2017 - November 2018, and December 2018 - November 2019. In each time frame, we code all months in each dataset past May as a placebo “treatment” month for all STAR-active precincts. If these estimates indicate a consistent drop in crime in the STAR precincts following May of each year, that would suggest the main results reflect seasonal patterns rather than the implementation of the STAR pilot. Instead, results based on these data consistently suggest no statistically detectable reductions in crime after June in any of

these prior years. We show these prior-year placebo estimates for static DD specifications (Table S6) and event studies (Figures S5-S7).

Several additional internal validity checks detailed in the SM Appendix provide evidence on the COVID-19 pandemic as a possible confound as well as related evidence that speaks more generally to the validity of the parallel-trends assumption. In particular, estimates near the bottom of Table S6 indicate the estimated effects of the STAR program are similar in specifications that only rely on data following the beginning of COVID shutdown orders (i.e., March 2020 through November 2020). Second, the estimated effects of the STAR program are also similar when based on generalized synthetic control (GSC) (31, 32) and comparative interrupted time series (CITS) designs (33) that explicitly accommodate the presence of pre-existing trends across treatment and comparison precincts (see Table S7 for results and SI text for details on these procedures).

Finally, the SI Appendix also presents analyses that explore the potential heterogeneity in these results. For example, the event-study results (Figure 3) suggest that the impact of the STAR program grew over time. However, estimates based on a semi-dynamic DD specification cannot reject the hypothesis that the impact of the STAR program is the same in each of the six months of operation ($p = 0.91$, Table S5). Additionally, DD specifications that allow for spatial spillover effects of the STAR program in geographically adjacent districts indicate that the estimated effect of STAR operations on neighboring precincts was small and statistically insignificant (Table S6). However, there do appear to be meaningful temporal spillover benefits of the STAR program within treated precincts. Specifically, the reductions in STAR-related crimes in treated precincts also occurred during days of the week and times when the program was not active (Table S6). This pattern is consistent with the hypothesis that the STAR program provided helpful services to individuals in crises that were somewhat persistent rather than brief and episodic. Furthermore, using the police categorization of crimes suggests that STAR operations led to a 14 percent reduction in overall crimes

(Table S8) and that these reductions were concentrated in three STAR-related categories (i.e., alcohol and drugs, disorderly conduct, and other crimes against persons).

Discussion

Police officers in the U.S. currently spend a substantial amount of their time responding to non-violent emergency calls for assistance, which often involve individuals experiencing mental-health or substance-abuse crises. However, police officers are not extensively trained to assist with such crises and most believe that such incidents are outside of their professional purview (34, 35). As a result, emergency calls for assistance may be engaged as criminal violations, sometimes with unnecessarily violent or even tragic consequences, when they can be better addressed as health issues. The widespread recognition of this issue has motivated initiatives to improve police training and cooperation with health professionals (e.g., Crisis Intervention Teams and Co-Response models). A less common but more dramatic innovation for responding to non-violent individuals in crisis is to delay or forego police involvement by sending a health-care team as first responders (i.e., a “Community Response” model). Though each of these programmatic models is grounded in a sensible theory of change, there is not currently credible, causal evidence on their effects (2, 13-16, 18).

In this study, we have presented the results of a pre-registered quasi-experimental design that examined the effects on crime of a community response program that dispatched a mental-health clinician and paramedic to non-violent emergency calls rather than first sending police. The Support Team Assistance Response (STAR) was a community-response program that operated as a pilot program for six months and provided service within eight police precincts in Denver’s central downtown area. Drawing on data of adult criminal incidents recorded regardless of whether the incident led to formal charges, from December 2019 through November 2020, we have employed a “difference-in-differences” (DD) model that

effectively compares the changes in criminal incidents both before and after the pilot program and across the treated and untreated precincts. We complement the results of this pre-registered design with a variety of robustness checks (e.g., alternative approaches to estimation and inference and falsification exercises based on prior years of data). We also examined whether STAR operations influenced the frequency of more serious offenses that were not directly targeted by the program and found no discernable impact.

We find that the program led to large and sustained reductions in STAR-related offenses in treated precincts, while unrelated offenses over the treatment period changed little in those same police precincts (Figure 2). Our comparative estimates suggest that the service reduced the number of STAR-related offenses in treated precincts by 34 percent over the six months of the pilot phase. While the average number of STAR-related offenses in our precinct-month sample is 34 (see Table S2), the frequency of these offenses in STAR-active precincts prior to treatment is much higher (i.e., averaging 84.3 offenses per precinct-month from December 2019 through May 2020). This impact estimate implies that the STAR pilot program prevented nearly 1,400 criminal offenses within the eight participating precincts and the six months of operation (i.e., $84.3 \times 0.34 \times 8 \times 6 = 1,376$). This program-induced reduction in offenses is broadly consistent with the scale of STAR operations. Specifically, the STAR team responded to 748 calls during our study window. At baseline, each STAR-related incident resulted in an average of 1.4 offenses in treated precincts. This suggests we should expect 748 field calls by STAR staff to result directly in just over 1,000 fewer offenses (i.e., 748×1.4). Our evidence that STAR operations also reduced targeted offenses during days and hours when the program staff were unavailable is consistent with the hypothesis that providing mental-health services reduces repeat offending by individuals in crisis. This finding also implies that the overall impact of the STAR program on criminal offenses is larger than the direct effect resulting from emergency calls to which they responded.

The evidence suggests that the STAR community-response program was effective in reducing criminal incidents and provides a compelling motivation for the continued implementation and assessment of this approach. However, successfully replicating the STAR program is likely to rely on key implementation details such as the recruitment and training of dispatchers and mental-health field staff as well as the successful coordination of their activities with the police. Furthermore, the generalizability of the community-response approach to a broader set of potentially preventable charges is uncertain and a design feature worthy of further study. Another important policy consideration is its cost effectiveness. The total cost of the six-month STAR pilot program was \$208,141 (36). One useful way to frame this public outlay is to note that the corresponding reduction of 1,376 offenses implies a cost of \$151 per offense reduced. To put this in perspective, the available estimates (3) suggest that the direct criminal-justice cost for a minor criminal offense (e.g., imprisonment and prosecuting) averages \$646 (in 2021 dollars). In other words, the direct costs of having police as the first responders to individuals in mental-health and substance-abuse crises are over four times as large as those associated with a Community-Response model. A fuller reckoning of the costs and benefits associated with Community Response models would also include the costs and benefits associated with any health care brokered by the first responders. Nonetheless, the results presented here suggest that Community-Response models merit careful consideration as a highly cost-effective way to reduce police engagement with non-violent individuals in crisis and to instead respond with appropriate health care.

Materials and Methods

We collected counts of criminal offenses using data made publicly available by the City and County of Denver, Colorado through their Open Data Catalog (ODC) and based on the National Incident Based Reporting System (NIBRS). There are differences between the data reported to NIBRS and what is reported on Denver's Open Data Catalogue that we use here, however. That is, NIBRS records all serious

incidents like homicide and arson, but only reports records of arrests made for less serious offenses like the STAR-related ones that are our focus here. Conversely, we have confirmed through correspondence with Dr. Matthew Lunn, the Manager of Strategic Initiatives for the Denver Police Department that their data record not only those crimes resulting in arrests or formal charges (i.e., ones involving subsequent prosecutorial decisions) but rather all criminal incidents recorded by the police. To that end, our main results identify program-induced reductions in substantiated criminal incidents identified by or reported to the police. Conceptually, these reductions combine the relabeling of existing behaviors that occur when individuals in crisis receive health care rather than being directed into the criminal-justice system and a reduction in criminal offenses by individuals in crisis who would offend repeatedly in the absence of health care.

The ODC contains incident-level data on all criminal offenses reported to law enforcement from January 2016 to November 2020. From that data catalogue, we constructed a panel dataset of criminal offenses observed in each of 36 precincts over each of 12 months for the period from December 2019 through November 2020 (i.e., 432 precinct-month observations). This period includes the six-month pilot phase and the six months prior to the pilot phase.

The single confirmatory hypothesis in our pre-registered analysis plan (<https://osf.io/3t8s7>) focuses on the impact of the STAR program in a static “difference-in-differences” (DD) specification that takes the following form:

$$Y_{pm} = \alpha_p + \gamma_m + \theta S_{pm} + \varepsilon_{pm} \quad (1)$$

where Y_{pm} is the natural log of STAR-related criminal offenses for precinct p in month m . The term, S_{pm} , is a binary indicator equal to 1 only for STAR-participating precincts observed during the period when the program was active. The coefficient of interest, θ , represents the effect of the STAR program conditional on fixed effects unique to each precinct and to each month (i.e., α_p and γ_m , respectively). The term, ε_{pm} , is a mean-zero error term with clustering at the precinct level. The static DD specification

in (1) embeds the assumption that the STAR program implies a one-time level shift in crimes. To explore possibly time-varying treatment effects, we also report the results of “semi-dynamic” DD that unrestrictively allow for effects to vary uniquely in each of the six treatment months. In the SM Appendix, we also present the results based on versions of equation (1) that use alternative approaches to estimation (e.g., Poisson and negative-binomial count-data specifications) and to inference (e.g., adjustments for finite-sample clustering bias, randomization inference).

This DD research design effectively compares the pre/post level of crimes in STAR-active precincts to the contemporaneous change in comparison precincts (i.e., those where STAR services were unavailable). A key identifying assumption of this design is that the time-varying changes within the comparison precincts provide a valid counterfactual for what would have happened in the treated districts in the absence of treatment. We examine the empirical validity of this assumption in two ways. One is to estimate “event study” DD specifications that unrestrictively allow for effects unique to treatment precincts in each month. The event-study estimates indicate the extent to which the treatment and comparison precincts had similar month-to-month variation in STAR-related crimes *before* the pilot began (see the SM Appendix for details). If treatment and comparison precincts have similar trends in STAR-related crimes in the months prior to STAR operations, it would be consistent with the internal validity of the DD design.

A second, important robustness check is to use equation (1) to estimate the impact of STAR operations on the more serious criminal offenses that are *not* directly related to STAR operations. If the estimates based on equation (1) are reliable, we would expect the DD design to indicate that the effect of STAR operations on STAR-unrelated crimes is comparatively small, if not indistinguishable from zero. However, if DD estimates indicate that STAR operations had large effects on crimes unrelated to STAR, it would suggest the existences of unobserved and confounding variables that are unique to the treated precincts in the treatment period. We formalize the idea

of using the crimes unrelated to STAR operations as a comparison condition that is unique to each precinct and month in “difference-in-difference-in-differences” (DDD) specifications. Specifically, we stack the precinct-by-month data for these two crime categories ($n = 864$) and estimate the following specification:

$$Y_{pom} = \alpha_{pm} + \gamma_{mo} + \delta_{po} + \theta S_{pom} + \varepsilon_{pom} \quad (2)$$

This specification includes unrestrictive fixed effects unique to each possible two-way interaction: precinct-month (α_{pm}), month-offense (γ_{mo}), and precinct-offense (δ_{po}). Critically, the DDD specification controls for unobserved determinants of crime unique to each precinct-month combination. The parameter of interest reflects the estimated effect associated with the three-way interaction unique to STAR-related offenses observed in treated precincts during the treatment period (i.e., S_{pom}).

References and Notes

1. S. R. Neusteter, M. O'Toole, M. Khogali, A. Rad, F. Wunschel, S. Scaffidi, M. Sinkewicz, M. Mapolski, P. DeGrandis, D. Bodah, H. Pineda, "Co-responding police-mental health programs: A review," Vera Institute of Justice, 2020.
2. A. Irwin, B. Pearl. "The community responder model," Center for American Progress, 2020.
3. M. A. Cohen, A. R. Piquero, New evidence on the monetary value of saving a high risk youth. *Journal of Quantitative Criminology* **25**, 25-49 (2009). doi: 10.1007/s10940-008-9057-3.
4. M. Jagannathan, "As activists call to defund the police, mental-health advocates say 'the time is now' to rethink public safety," *MarketWatch*, 19 June 2020.
5. US Department of Health & Human Services, "Mental health myths and facts," <https://www.mentalhealth.gov/basics/mental-health-myths-facts> [accessed 12 March 2021].
6. M. E. Rueve, R. S. Welton, Violence and mental illness. *Psychiatry* **5**, 34-48 (2008).
7. IACP Law Enforcement Policy Center, "Responding to persons experiencing a mental health crisis," 2018.
8. D. A. Fuller, H. R. Lamb, M. Biasotti, J. Snook, "Overlooked in the undercounted: The role of mental illness in fatal law enforcement encounters," Treatment Advocacy Center, 2015.
9. B. Dahlberg, "Rochester Hospital released Daniel Prude hours before fatal encounter with police," *NPR*, 29 September 2020.
10. T. Elfrink, "'He's a small child': Utah police shot a 13-year-old boy with autism after his mother called 911 for help," *The Washington Post*, 8 September 2020.

11. NBC10 Staff, "Walter Wallace Jr. struggled with mental health issues, family says," *NBC 10 Philadelphia*, 29 October 2020.
12. M. T. Compton, M. Bahora, A. C. Watson, J. R. Oliva, A comprehensive review of extant research on crisis intervention team (CIT) programs. *Journal of the American Academy of Psychiatry and the Law Online* **36**, 47-55 (2008).
13. G. K. Shapiro, Co-responding police-mental-health programs: A review. *Administration and Policy in Mental Health and Mental Health Services Research* **42**, 606-620 (2015). doi: 10.1007/s10488-014-0594-9.
14. S. Puntis, D. Perfect, A. Kirubarajan, S. Bolton, F. Davies, A. Hayes, E. Harriss, A. Molodynski, A systematic review of co-responder models of police mental health 'street' triage. *BMC Psychiatry* **18**, 1-11 (2018). doi: 10.1186/s12888-018-1836-2.
15. J. Peterson, J. Densley, Is crisis intervention team (CIT) training evidence-based practice? A systematic review. *Journal of Crime and Justice* **41**, 521-534 (2018). doi: 10.1080/0735648X.2018.1484303.
16. S. A. Taheri, Do crisis intervention teams reduce arrests and improve officer safety? A systematic review and meta-analysis. *Criminal Justice Policy Review* **27**, 76-96 (2016). doi: 10.1177/0887403414556289.
17. C. Seo, B. Kim, N. E. Kruis, Variation across police response models for handling encounters with people with mental illnesses: A systematic review and meta-analysis. *Journal of Criminal Justice* **72**, 101752 (2020). doi: 10.1016/j.jcrimjus.2020.101752.
18. C. Seo, B. Kim, N. E. Kruis, A meta-analysis of police response models for handling people with mental illnesses: Cross-country evidence on the effectiveness. *International Criminal Justice Review*, 1-21 (2020). doi: 10.1177/1057567720979184.
19. T. Richards, "AOC's defund-the-police pick for NYC mayor will 'embolden the bad guys,' Kerik warns." *The Washington Examiner*. 8 June 2021.

20. J. Stepman, “Violent Crime Keeps Surging as More on Left Admit It’s Foolish to ‘Defund Police’,” *The Daily Signal*, 3 June 2021.
21. Z. Elinson, D. Frosch, J. Jamerson, “Cities reverse defunding the police amid rising crime,” *The Wall Street Journal*, 26 May 2021.22. M. McConnell, “Democrats’ ‘defund the police’ efforts coincide with explosion in violent crime,” *Senate Republican Leader*, 27 May 2021.
22. White Bird Clinic, “Crisis Assistance Helping Out On The Streets.” 2018. <https://www.mentalhealthportland.org/wp-content/uploads/2019/05/2018CAHOOTSBROCHURE.pdf>. [accessed 12 March 2021].
23. Denver Office of Economic Development, “Gentrification study: Mitigating involuntary displacement,” 2016.
24. C. Herring, D. Yarbrough, L. M. Alatorre, Pervasive penalty: How the criminalization of poverty perpetuates homelessness. *Social Problems* **67**, 131-149 (2020). doi: 10.1093/socpro/spz004.
25. T. Gowan, *Hobos, Hustlers, and Backsliders: Homeless in San Francisco*. (U of Minnesota Press, 2010).
26. T. Gowan, The nexus: Homelessness and incarceration in two American cities. *Ethnography* **3**, 500-534 (2002). doi: 10.1177/1466138102003004007.
27. D. Mitchell, *The Right to the City: Social Justice and the Fight for Public Space*. (Guilford Press, New York, 2003).
28. Denver Justice Project, “Press release: DJP helps launch alternative public health emergency response pilot in Denver,” 2020.
29. STAR Evaluation Team, “STAR program evaluation,” 2021. https://wp-denverite.s3.amazonaws.com/wp-content/uploads/sites/4/2021/02/STAR_Pilot_6_Month_Evaluation_FINAL-REPORT.pdf [accessed 12 March 2021].

30. J. A. Hausman, B. H. Hall, Z. Griliches, Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* **52**, 909-938 (1984). doi: 10.2307/1911191.
31. Y. Xu, Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Anal.* **25**, 57-76 (2017). doi:10.1017/pan.2016.2.
32. L. Liu, Y. Wang, Y. Xu, A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. SSRN: <http://dx.doi.org/10.2139/ssrn.3555463>, 2020.
33. W. R. Shadish, T. D. Cook, D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Houghton Mifflin, Boston, 2002).
34. J. Peterson, J. Densley, G. Erickson, Evaluation of 'The R-Model' crisis intervention de-escalation training for law enforcement. *The Police Journal* **93**, 271-289 (2020). doi: 10.1177/0032258X19864997.
35. J. Hails, R. Borum, Police training and specialized approaches to respond to people with mental illnesses. *Crime & Delinquency* **49**, 52-61 (2003). doi: 10.1177/0011128702239235.
36. E. Schmelzer, "Call police for a woman who is changing clothes in an alley? A new program in Denver sends mental health professionals instead," *The Denver Post*, 6 September 2020.
37. J. E. Pustejovsky, E. Tipton, Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics* **36**, 672-683 (2018). doi: 10.1080/07350015.2016.1247004.

Acknowledgments

We thank Jeremy Pyne for generating the maps presented in this paper, Dagoberto Cortez for guidance on relevant urban gentrification literature, Chris Richardson and Carleigh Sillon for helping us better understand the STAR program Denver, CO., and Matthew Lunn for helping us better understand how the City and County of Denver record and publicly report their incident-level crime data.

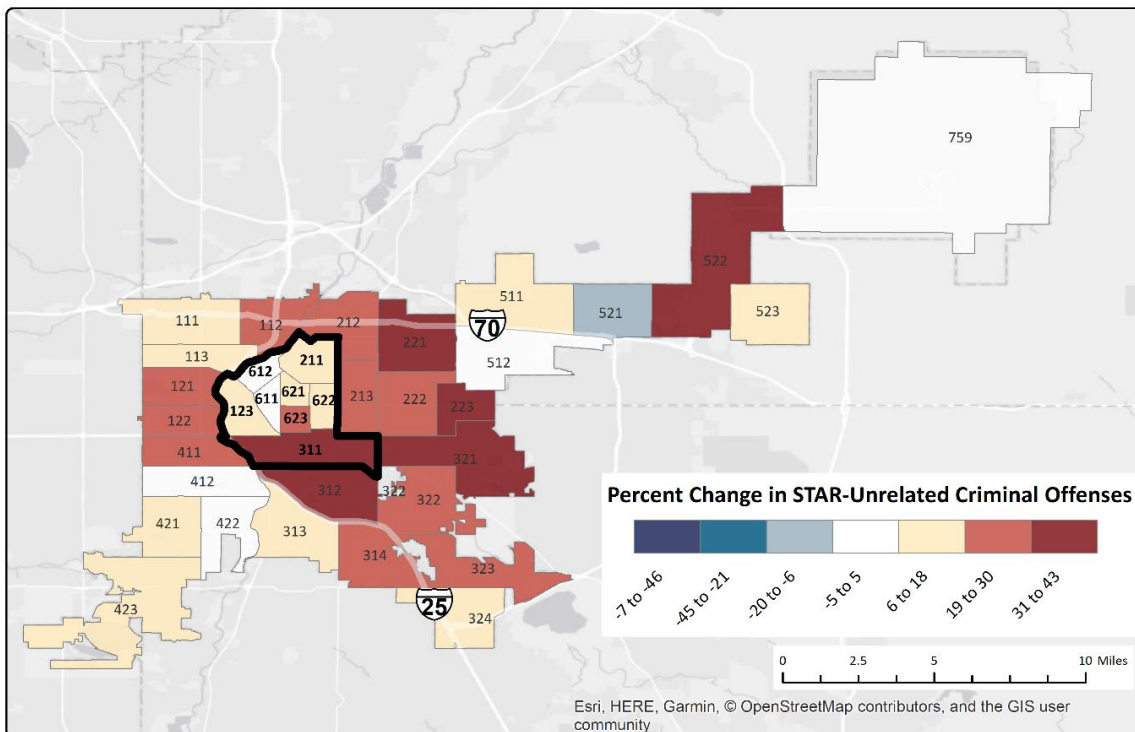
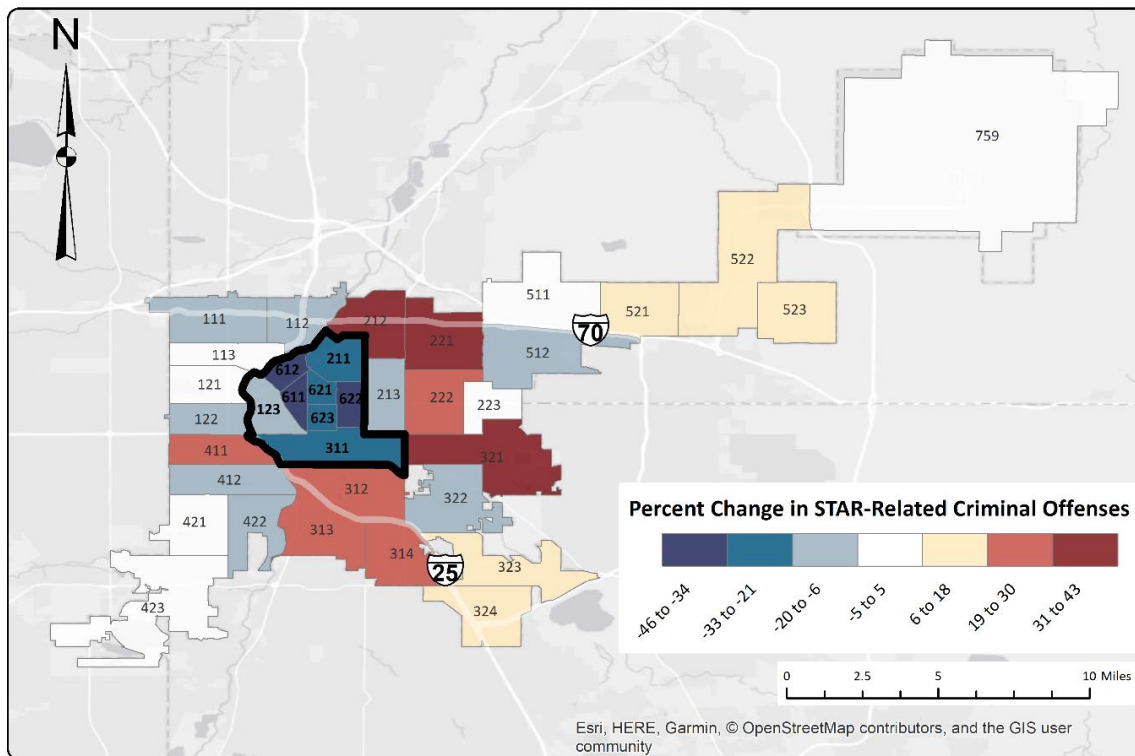


Fig. 1. Changes in criminal offenses before and after STAR pilot implementation. The thick black line surrounds the police precincts where the STAR program was active.

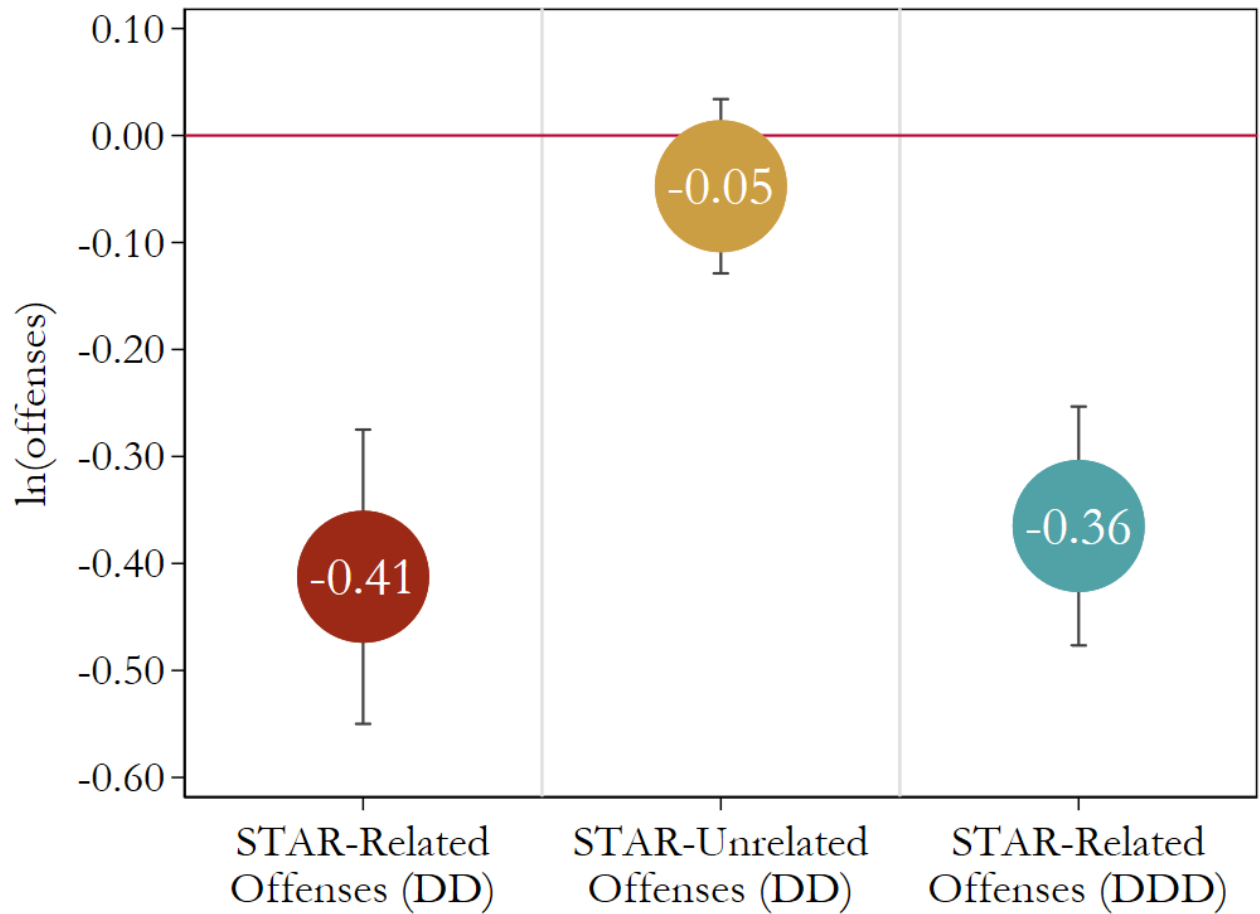


Fig. 2. Estimated effects of the STAR program on criminal offenses. The difference-in-differences (DD) estimates are based on 432 precinct-month observations and condition on precinct fixed effects and month fixed effects. The difference-in-difference-in-differences (DDD) estimates are based on the stacked precinct-month data for STAR and non-STAR offenses ($n = 864$). The DDD estimates condition on fixed effects unique to each category of the following 2-way interactions: precinct-by-month, precinct-by-STAR offense, and month-by-STAR offense. The outcome variables are the natural log of the offense counts. Dots are coefficients; bars are 95% confidence intervals. See Table S4 for numerical results.

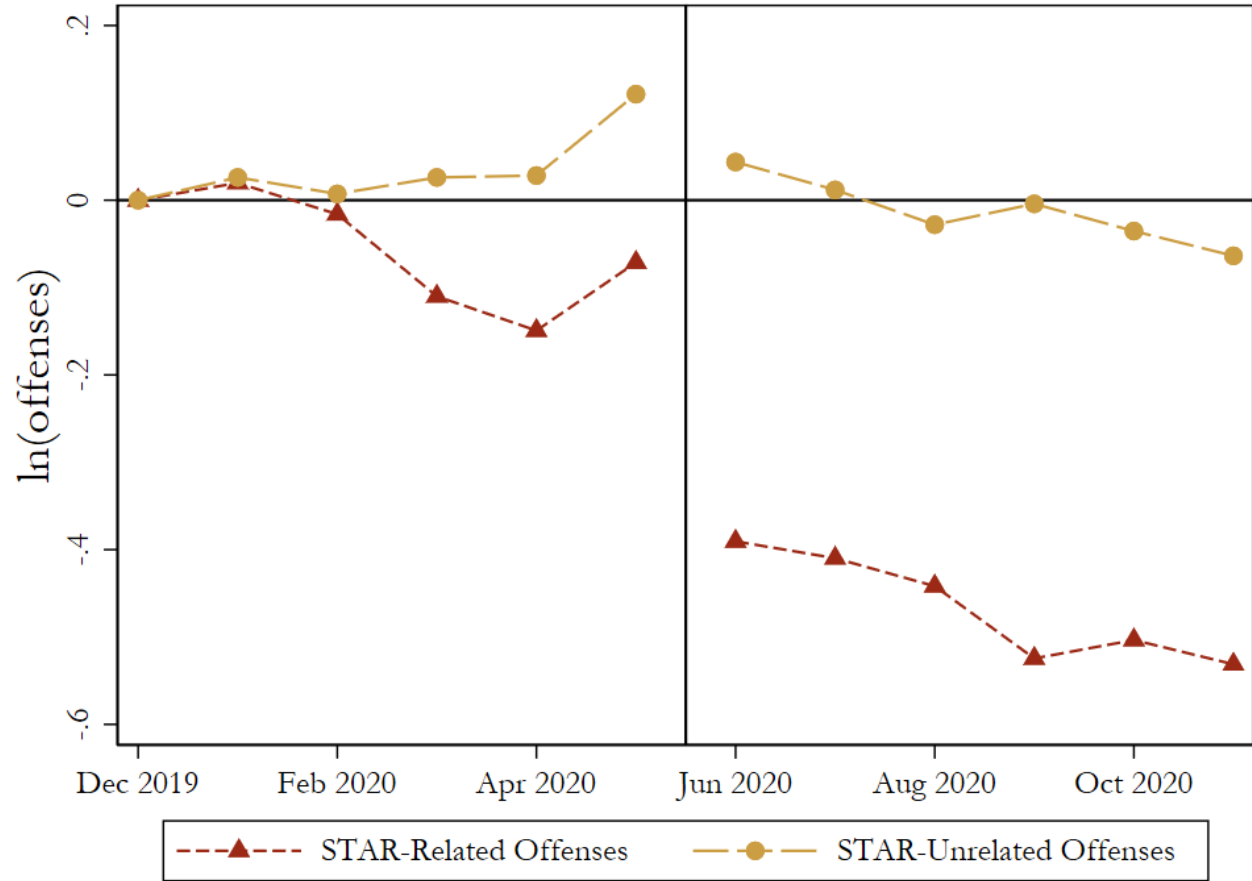


Fig. 3. Event study model. The difference-in-differences event-study (DD) estimates are based on 432 precinct-month observations and condition on precinct fixed effects and month fixed effects. The outcome variables are the natural log of the STAR-related and STAR-unrelated offense counts. The event-study estimates identify for each outcome the regression-adjusted treatment-comparison differences by month relative to the first time period (i.e., December 2019). The vertical line separates pre-treatment months from the months after the STAR pilot program began in June, 2020. See Table S5 for numerical results.

Supplementary Materials

Materials and Methods

City and Police Precinct Characteristics

Figure S1 maps the police districts and neighborhoods in which the STAR program was active. The STAR pilot program operated in select police precincts, “from York St. to I-25 east to west and 38th St. / 40th Ave. to 6th Avenue north to south) and along the South Broadway corridor to Mississippi Ave., with service also being provided to the temporary shelters at the Denver Coliseum and National Western Complex” (25). This constitutes mainly the central downtown Denver area and includes police precincts 123, 211, 311, 611, 612, 621, 622, and 623 (see Figure S1). Table S1 presents descriptive demographic and socioeconomic data on these neighborhoods as well as for the entire city based on data from the American Community Survey (ACS). Denver, CO is a city with a population of 678,467. In 2017, the median household income was \$59,179 (15% poverty). Just under half (46%) had a bachelor’s degree or higher, and the same proportion (46%) are people of color.

Sample Traits

Table S2 shows the categorization of specific offenses in the City’s NIBRS-based data as STAR-related or unrelated. As noted in the main text, these categorizations were based on two independent ratings (91% agreement; kappa = 0.73) with investigation and reconciliation of differences completed prior to preregistration and analysis. Table S3 presents key descriptive statistics based on the precinct-month analytical sample ($n = 432$). Specifically, this table shows the means for STAR-related and unrelated offenses (i.e., 33.7 and 156.0, respectively). It also shows the offenses across 10 broad and mutually exclusive categories of offenses used in the City’s data file. Five uncommon categories of offenses unrelated to the STAR program’s mission (i.e., arson, murder, robbery, sexual assault, white-collar offenses) are excluded.

Offense Coding

The incident data identify 199 types of offenses organized into 15 broader and mutually exclusive categories. For our primary analyses, we categorized each recorded offense by whether it was directly related to STAR services. Specifically, prior to our pre-registered analysis, two independent coders rated the categorization of each offense type. Raters had 91 percent agreement on offense type codes (kappa = 0.73). Coders met and reconciled remaining discrepancies. The offenses identified as STAR-relevant include trespassing, disturbing the peace, possession of illegal drugs, indecent exposure, alcohol violations, loitering, failure to obey police orders, police interference, and public disorder. Prior to STAR operations, the offenses identified as STAR-relevant offenses constituted 20 percent of the offenses reported by Denver police. As a complement to this binary categorization of offenses, we also show results based on broad, mutually exclusive categories the City reports.

Treatment Heterogeneity and Evidence of Robustness

The pre-registered “static” DD specification represented in equation 1 (see main text) assumes that the treatment effect is constant over time. However, the effects of the STAR program could instead have dynamic features. To test for time-varying treatment effects, we also employ a semi-dynamic DD model that unrestrictively allows for treatment effects unique to the month immediately after a precinct first participates and up to five months later:

$$Y_{pm} = \alpha_p + \gamma_m + \sum_{n=0}^5 \delta_{-n} S_{p,m-n} + \varepsilon_{pm}$$

In this model, the three coefficients of interest are represented by δ_n , which identify the effects of STAR in the first month of the program (i.e., $S_{p,m-0}$) as well as the current effect of having begun one month earlier (i.e., $S_{p,m-1}$), two months earlier (i.e., $S_{p,m-2}$), and so on. We then test the equivalence of these coefficients of interest using the null hypothesis of a constant treatment effect:

$$H_0: \delta_0 = \delta_{-1} = \delta_{-2} = \delta_{-3} = \delta_{-4} = \delta_{-5}$$

We report the semi-dynamic results, both for DD and DDD specifications, in Table S4. Hypothesis tests consistently fail to reject the null hypothesis of a common treatment effect across the 6-month pilot period.

Table S5 reports DD estimates of the impact of the STAR program on overall offenses and on offenses across the broad and mutually exclusive categories defined in the City’s NIBRS-based data. The point estimates indicate that the STAR program reduced the natural log of total offenses by a statistically significant 0.15, which implies the 14 percent reduction noted in the main text (i.e., $e^{-0.15} - 1$). The estimates by category indicate that these reductions were plausibly concentrated in offenses such as “Alcohol and drugs” (i.e., -0.53), “disorderly conduct” (i.e., -0.20), and “other crimes against people” (i.e., -0.14).

Table S6 presents the key results from a variety of alternative specifications that probe the robustness and heterogeneity of the confirmatory finding. First, we consider alternative approaches to conducting inference in this application. Our main estimates allow for precinct-specific clustering in the error term associated with criminal offenses that is heteroscedastic-consistent. However, because there are only 36 unique precincts, this clustering approach may be subject to finite-sample biases. To examine this concern, we report the results based on the procedure recently introduced by Pustejovsky and Tipton (37). The results are quite similar to our reported findings.

As a further and unrestrictive check on our main inference, we also conducted randomization inference with respect to the confirmatory finding. Specifically, over 100,000 replications, we randomly assigned treatment status within precincts and estimated the “impact” of the STAR program. Randomization inference has a particular appeal in applications like this because the data may be better understood as having “design-based” variation in what units are treated rather than having variation due to being drawn from a larger hypothesized population. Figure S2 shows the histogram of estimated effects based on this permutation procedure. Because treatment status was assigned randomly, this distribution can be understood as the distribution of treatment effects when the null hypothesis of no effect is true. Over the 100,000 replications, none of the estimates in this distribution was as large in absolute value as the estimate based on the actual data (i.e., -0.41). This implies a randomization-inference p -value that is less than 0.00001.

Table S6 also presents results based on alternative estimation procedures and constructions of the analytical sample. Specifically, Table S6 presents the conditional maximum likelihood (CML) estimates of Poisson and negative binomial specifications that explicitly recognize both the count nature of the offense data and the presence of fixed effects (28). The resulting estimates are quite similar to those based on the pre-registered DD specification. Table S6 also presents the main DD results when dropping data from a STAR-participating police precinct (i.e., precinct 311) where program activities were targeted to a main corridor rather than intending to be active precinct-wide. Though there is no clear reason to expect biases from the onset of the COVID-19 pandemic, especially conditional on month fixed effects, Table S6 also shows the results of using data only from March 2020 (i.e., the onset of the shutdown) onward. Both data edits result in DD estimates consistent with our main finding.

Finally, Table S6 also presents the results of exploring two particular forms of treatment heterogeneity. First, we explored the possibility that the STAR program also led to crime reductions in geographically adjacent precincts. Specifically, we created an additional treatment indicator equal to one only for precincts that were adjacent to STAR precincts when the STAR program was active. The estimated effect reported in Table S6 indicates that we cannot reject the null hypothesis of no effect in neighboring precincts. We also explore possibly heterogeneous treatment effects across days of the week and times of the day when the STAR program was active. As the main text notes, the program was only active Monday through Friday, 10AM to 6PM. We created separate counts for STAR-related offenses that occurred within and outside these weekly windows. The results in Table S6 indicate that the STAR program led to similar reductions in targeted offenses across both time periods. As noted in the main text, this finding is consistent with the hypothesis that the program brought into the health-care system individuals in crisis who would otherwise commit police-reported offenses at other times of the week (i.e., evenings and weekends) as well.

The main text underscores two types of evidence consistent with the internal validity of the pre-registered DD results. One is the absence of a meaningful impact on offenses rated as unrelated to STAR prior to the analysis (i.e., column 3 in Table S4). Table S7 presents the results of a second and important type of evidence. A central and maintained identifying assumption of our pre-registered DD approach is that the month-to-month outcome changes among comparison precincts (i.e., those without a change in treatment status) provide a valid counterfactual for what would have changed for treatment precincts in the absence of treatment. This “parallel trends” assumption is fundamentally untestable. However, we can provide empirical evidence on the validity of this important assumption through unrestrictive “event study” specifications that allow us to examine whether treatment and comparison group precincts had similar month-to-month changes in outcomes *prior* to the onset of treatment. To the extent that this hypothesis is true, it is consistent with the parallel-trends assumption. We examine this question through event-study specifications of the following form:

$$Y_{pm} = \alpha_p + \gamma_m + \sum_{\tau=1}^5 \delta_{\tau} S_{p,m+\tau} + \sum_{n=0}^5 \delta_{-n} S_{p,m-n} + \varepsilon_{pm}$$

This event-study specification effectively extends the semi-dynamic specification (equation 2) to allow for fixed effects unique to each month prior to participating in STAR (i.e., “leads” of treatment adoption). That means the coefficients of interest are represented as δ_{-n} and δ_{τ} , which designate the “effect” for precinct p in month m of participation in STAR n months in the future or τ months in the past. The reference category includes those never participating in STAR and those in six months prior to their first participation in STAR. To examine the assumption of parallel trends, we test whether, *prior* to their participation in STAR, treatment precincts have month-to-month changes in outcomes distinct from comparison precincts:

$$H_0: \delta_5 = \delta_4 = \delta_3 = \delta_2 = \delta_1 = 0$$

We report the event-study results, both for STAR-related and unrelated offenses, in Table S7 and Figure 3 in the main text. The results are consistent with the parallel-trends assumption, indicating that we cannot reject the null hypothesis that the treated precincts had month-to-month changes similar to the comparison districts in the months *prior* to the program activity.

Pre-Registration Plan

The following is our detailed pre-registration plan (<https://osf.io/3t8s7>), filed on February 14, 2021 prior to any data analysis related to the study.

A. Study Information

1. *Hypotheses*

Precincts participating in the STAR program will have reduced prevalence of criminal offenses related to mental health, poverty, homelessness, and substance abuse in the City of Denver.

- B. Design Plan

1. Study type

Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, “natural experiments,” and regression discontinuity designs.

2. *Blinding*

No blinding is involved in this study.

3. *Study design*

We focus on recorded offenses in each city precinct in a given month from December 1, 2019 through November 30, 2020. This time period represents the six-month pilot phase of STAR (June 2020-November 2020) and the six months prior to the pilot beginning. This design strategy allows us to take advantage of our panel dataset in months surrounding implementation. Our analytical sample consists of 36 precincts and 432 precinct-month observations, from December 2019 through November 2020.

- C. Sampling Plan

1. *Existing Data*

Registration prior to analysis of the data

2. *Explanation of existing data*

The data come from open access police records provided by the city of Denver, CO (<https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-crime>). These data include criminal incident records from January 2, 2016 through January 15, 2021 involving adults. Due to legal restrictions, these data do not report crimes that by nature involve juveniles as victims (e.g., child abuse offenses), suspects or witnesses. These data also exclude “unfounded” incidents, which authorities have determined did not actually occur after they are reported.

3. *Data collection procedures*

We downloaded open access police records provided by the city of Denver, CO (<https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-crime>). We

retain recorded offenses in each city precinct in a given month from December 1, 2019 through November 30, 2020. This time period represents the six-month pilot phase of STAR (June 2020–November 2020) and the six months prior to the pilot beginning.

4. *Sample Size*

Our analytical sample consists of 36 precincts and 432 precinct-month observations, from December 2019 through November 2020.

5. *Sample size rationale*

This sampling allows for observation of criminal offenses in the city six months before and six months after the beginning of the STAR program, which allows for ample observation of pre and post treatment outcomes, tests of critical model assumptions, and for dynamic effects of the program.

D. Variables

1. *Measured variables*

Our outcome of interest is semi-logged precinct-month counts of STAR-related types of criminal offenses. The City of Denver codes recorded criminal offenses into fifteen overarching categories, including aggravated assault, arson, auto theft, burglary, drug and alcohol offenses, larceny, murder, public disorder, robbery, sexual assault, theft from motor vehicles, traffic accidents, white collar crimes, other crimes against individuals, and all other crimes. These categories give some sense of the types of crimes that might be related to the STAR programs aims but continue to carry a substantial amount of noise for our treatment estimates. From those 15 categories, offenses are differentiated by 199 different types. We coupled information on offense type descriptions and data on their frequencies with independent rater coding to identify those offenses that are STAR-relevant and those that are not. We measure treatment status by capturing each precinct's monthly participation in the STAR program. Using these data, we construct a simple binary indicator equal to 1 for precinct-month observations from precincts who participate in STAR during a given month (i.e., a "static" measure of treatment). We also use the timing of STAR participation to define less restrictive and flexibly dynamic measures of program participation. These include binary indicators for the month that the program began (June 2020) and separate indicators for being one through five months after that first participation month. These measures flexibly allow for the initial participation in STAR to have effects that increase or decline over time.

2. *Indices*

From the City of Denver's 15 overarching criminal offense categories, we differentiate those offenses that are most related to the types of calls that the STAR team will respond to from other types of offenses that are unlikely to either substitute for a noncriminal STAR team visit or would result from an escalation of such non-criminal offenses. From those 15 offense categories, offenses are differentiated by 199 different types. We coupled information on offense type descriptions and data on their frequencies with independent rater coding to identify those offenses that are STAR-relevant and those that are not.

E. Analysis Plan

1. *Statistical models*

Our main confirmatory analysis is based on a difference-in-differences (DD) design, which assumes that STAR activity in a given precinct and month leads to a constant, one-time change in STAR-related criminal offenses for participating precincts. We do so by comparing changes in these outcomes among precincts participating in STAR to outcomes of precincts that either never participated or had yet to participate in STAR. The outcome will be a semi-logged count of STAR-related criminal offenses. The predictors will be (1) an indicator of a treated precinct in a treated month, (2) precinct fixed effects, and (3) month fixed effects. Standard errors will be clustered at the precinct level.

2. *Transformations*

We use precinct-month counts of STAR-related criminal offenses in the panel data to estimate the effects of the STAR program on the number of offenses committed in each precinct. We transform the outcome variable, which is the semi-logged count of STAR-related criminal offenses for precinct p in month m .

3. *Inference criteria*

We will make inferences of our confirmatory analysis using two-tailed tests and p-values of $p < .10$. We will report p-values differently based on thresholds of $p < .01$, $p < .05$, and $p < .10$.

4. *Data exclusion*

We exclude data prior to December 2019 and after November 2020.

5. *Missing data*

There are no instances of missing precinct-month data, including criminal offenses recorded in a given precinct-month. Thus, we observe no missing data for our confirmatory analyses. However, in some exploratory analyses we examine program effects at the precinct-week level. For instances in which there are no STAR-related offenses in a given week, we will replace the missing value with the natural log of 0.5. We do the same for STAR-unrelated offenses in exploratory analyses.

6. *Exploratory analysis*

We will conduct a number of exploratory analyses. First, to test for time-varying treatment effects, we next employ a semi-dynamic DD model that unrestrictively allows for treatment effects unique to the month immediately after a precinct first participates and up to seven months later. We then test the equivalence of these coefficients of interest using the null hypothesis of a constant treatment effect. Second, we will conduct an "event study" analysis. A crucial maintained assumption of our DD approach is that the month-to-month outcome changes among "control" precincts (i.e., those without a change in treatment status) provide

a valid counterfactual for what would have changed for treatment precincts in the absence of treatment. This “parallel trends” assumption is fundamentally untestable. However, we can provide qualified evidence on the validity of this important assumption through unrestrictive “event study” specifications that allow us to examine whether treatment and control group precincts had similar month-to-month changes in outcomes prior to the onset of treatment. To the extent that this hypothesis is true, it is consistent with the parallel trends assumption. We examine this question through event-study specifications. Third, because these data also include counts of criminal offenses that are unrelated to the STAR programs goals, there is an opportunity to test a “triple diff” (DDD) research design that allows us to account for unobserved disturbances in precinct-month observations. Stacking our data at the precinct-month-(STAR & non-STAR) offense level, the DDD specification includes fixed effects for all two-way interactions. Fourth, we will rerun the static DD model for each of the 15 criminal offense category outcomes reported in the original dataset. Fifth, to test for potential differential effects of the COVID pandemic on criminal offenses, we rerun the confirmatory analysis but only include offenses from March 2020 through November 2020. Sixth, we rerun the confirmatory analysis using a count outcome in a negative binomial precinct fixed effects model. Seventh, we analyze the confirmatory outcome during STAR-eligible and STAR-ineligible times. Eighth, another model tests for spillover effects of the STAR program in precincts adjacent to the participating precincts. Ninth, we examine static and semi-dynamic program effects at the precinct-week level, for STAR-related and STAR-unrelated criminal offenses.

Deviations from Pre-Registration Plan

Our main results do not deviate from the pre-registration plan. However, we have added several additional exploratory analyses not reported in the original pre-registration plan. First, in a robustness check we recode “simple assaults”, “simple assaults on police officers”, and “disarming a piece officer” as STAR-unrelated offenses (see Table S6). Second, we include a static DD model in which we recode May 2020 as a “treatment” month among STAR-active precincts, to test for anticipation to the program’s start (see Table S6). Third, in Table S7 we include an additional pre-trends F-test for only months during COVID restrictions (i.e., March 2020 – May 2020). Fourth, we conduct placebo static DD and event study tests in years prior to our study window (see Table S6 and Figures S5-S7). Finally, we conduct additional robustness checks using a generalized synthetic control (GSC) design and a comparative interrupted time series (CITS) approach (see Table S8).

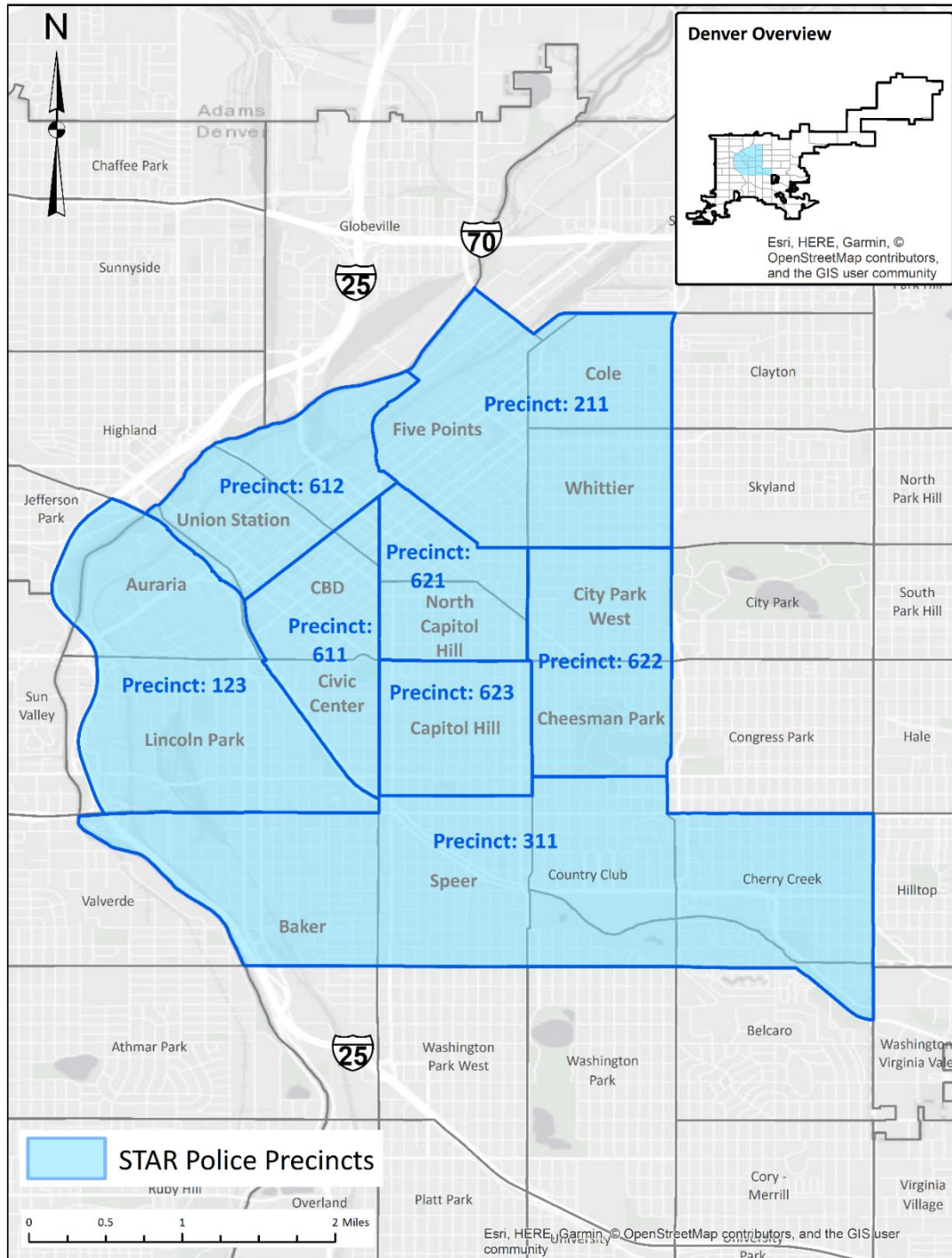


Figure S1. STAR police precincts and neighborhoods. Significantly affected police precincts and neighborhoods are bolded.

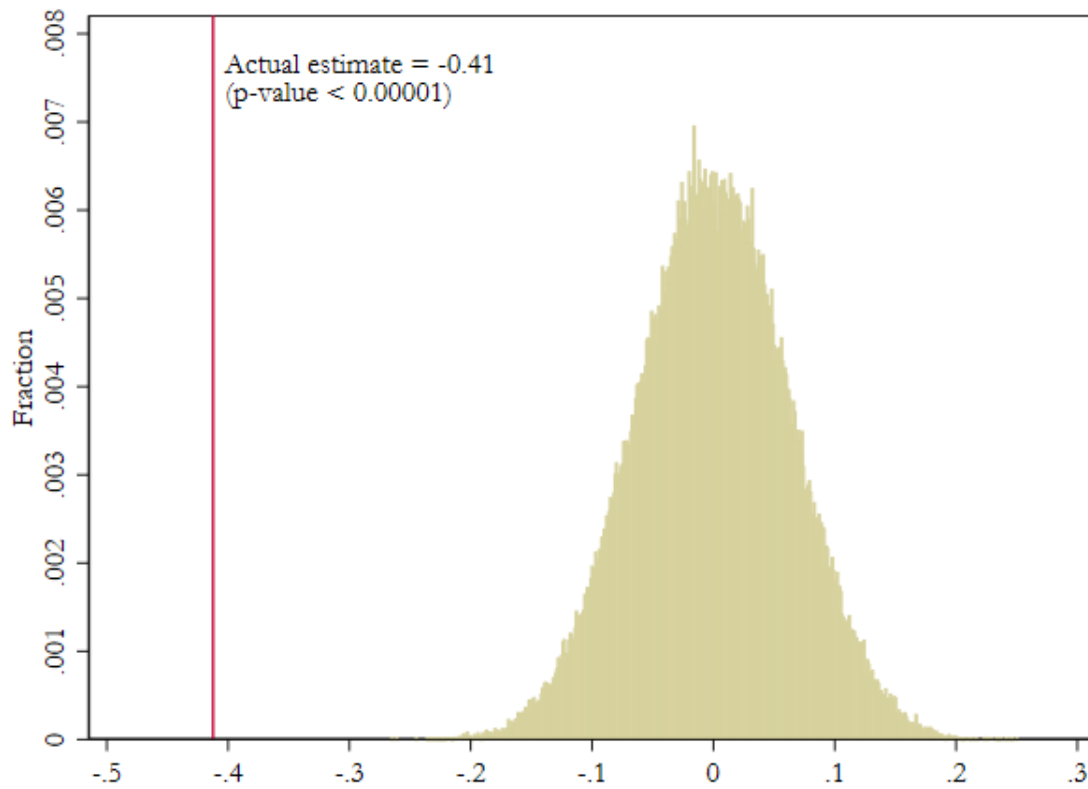


Figure S2. Null Distribution of Estimated Effects (100,000 Replications). Distribution represents results from 100,000 randomized simulations of the data under the assumption that the effects observed from the data were generated at random. The red vertical line represents the actual observed effect size from the data.

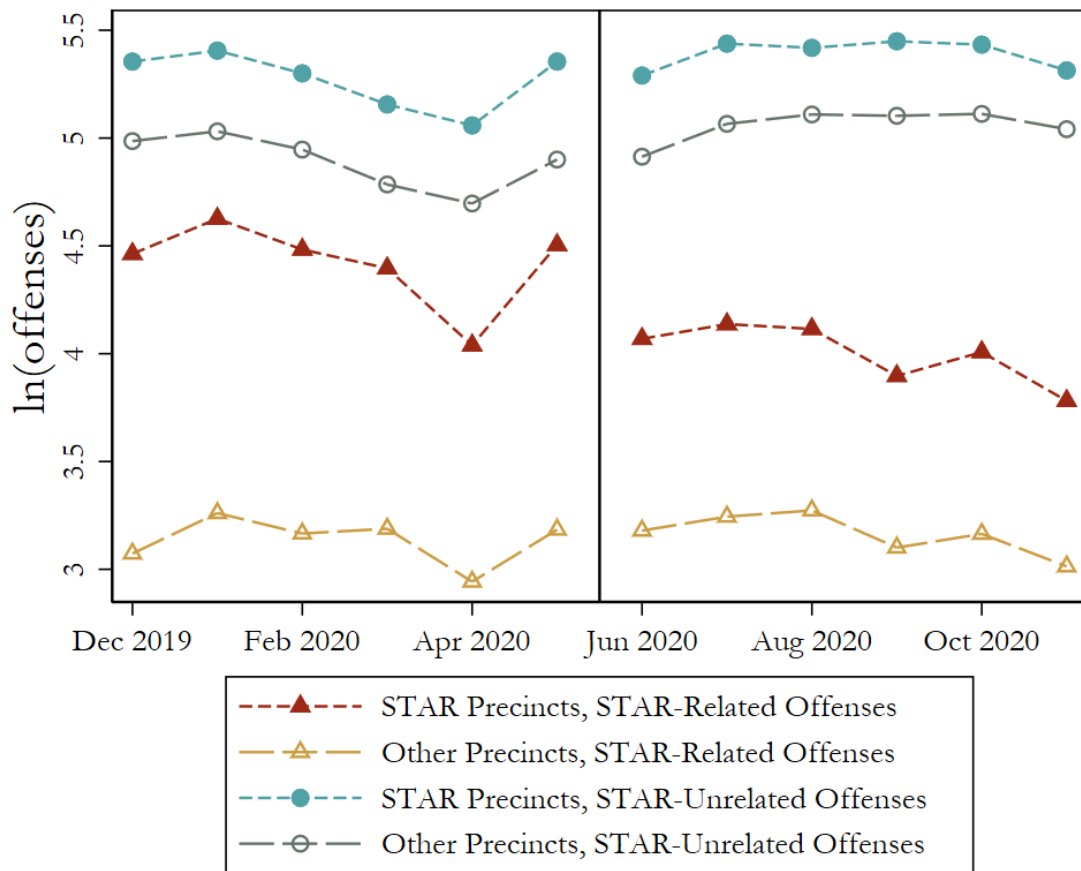


Figure S3. Descriptive Monthly Trends in Offense Types. Conditional means of offenses are based on 432 precinct-month observations of all Denver police precincts, from December 2019 through November 2020. The outcome variables are the natural log of the offense counts, differentiated by STAR and non-STAR precincts as well as by STAR-related and unrelated offenses. The vertical line separates months before and after the STAR pilot program began.

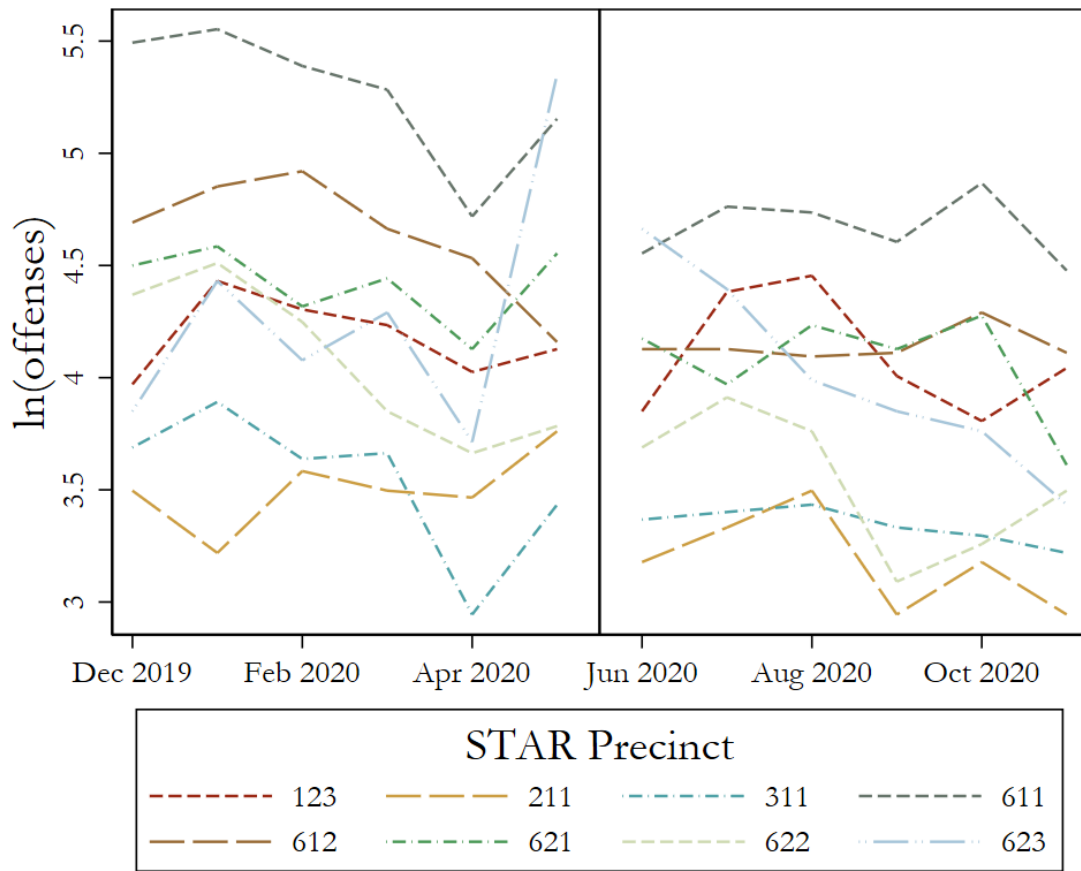


Figure S4. Descriptive Monthly Trends in STAR-Related Offenses among STAR Precincts. Lines are conditional means of STAR-related offenses based on 88 precinct-month observations from December 2019 through November 2020—one for each of the eight STAR-active precincts. The outcome variables are the natural log of the STAR-relevant offense counts. The vertical line separates months before and after the STAR pilot program began.

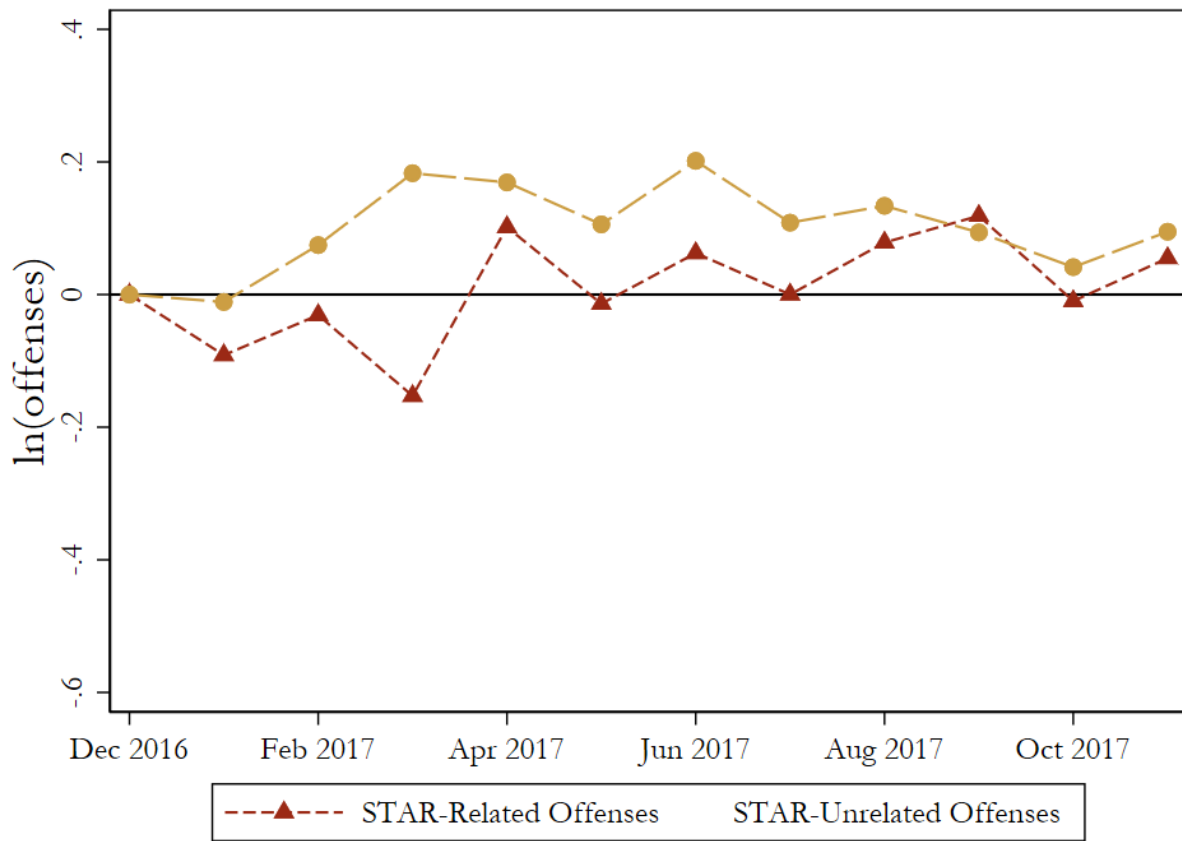


Figure S5. Placebo Event Study (Dec. 2016 – Nov. 2017). This placebo check mirrors the analyses represented in Figure 3 and Table S7, but applied to a time period other than our study window. Specifically, the difference-in-differences (DD) estimates are based on 432 precinct-month observations from December 2016 through November 2017 and condition on precinct fixed effects and month fixed effects. The outcome variables are the natural log of offense counts, differentiated by those that are STAR-related and those that are not. The horizontal line at zero denotes the baseline levels of offenses. The vertical line separates months of the year when the STAR program would and would not have been effective if the STAR program had begun in this time window rather than on June 2020.

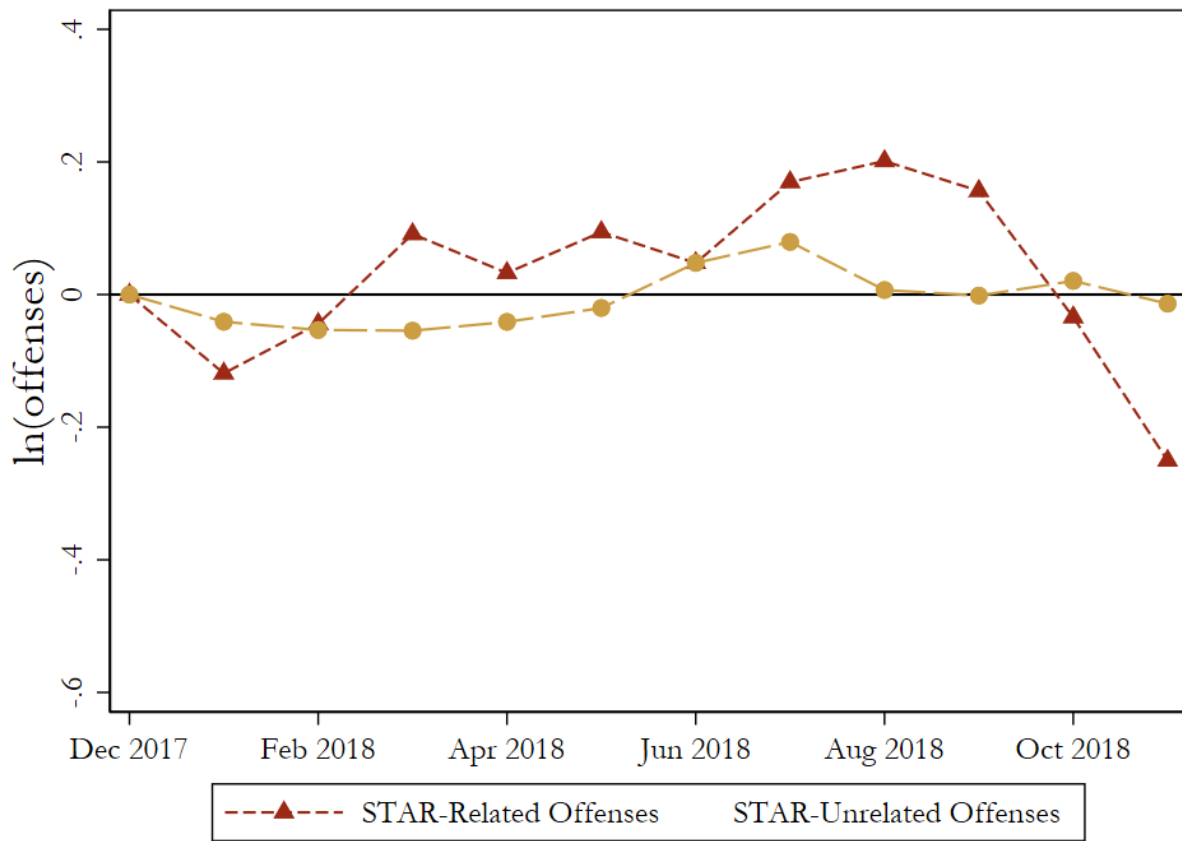


Figure S6. Placebo Event Study (Dec. 2017 – Nov. 2018). This placebo check mirrors the analyses represented in Figure 3 and Table S7, but applied to a time period other than our study window. Specifically, the difference-in-differences (DD) estimates are based on 432 precinct-month observations from December 2017 through November 2018 and condition on precinct fixed effects and month fixed effects. The outcome variables are the natural log of offense counts, differentiated by those that are STAR-related and those that are not. The horizontal line at zero denotes the baseline levels of offenses. The vertical line separates months of the year when the STAR program would and would not have been effective if the STAR program had begun in this time window rather than on June 2020.

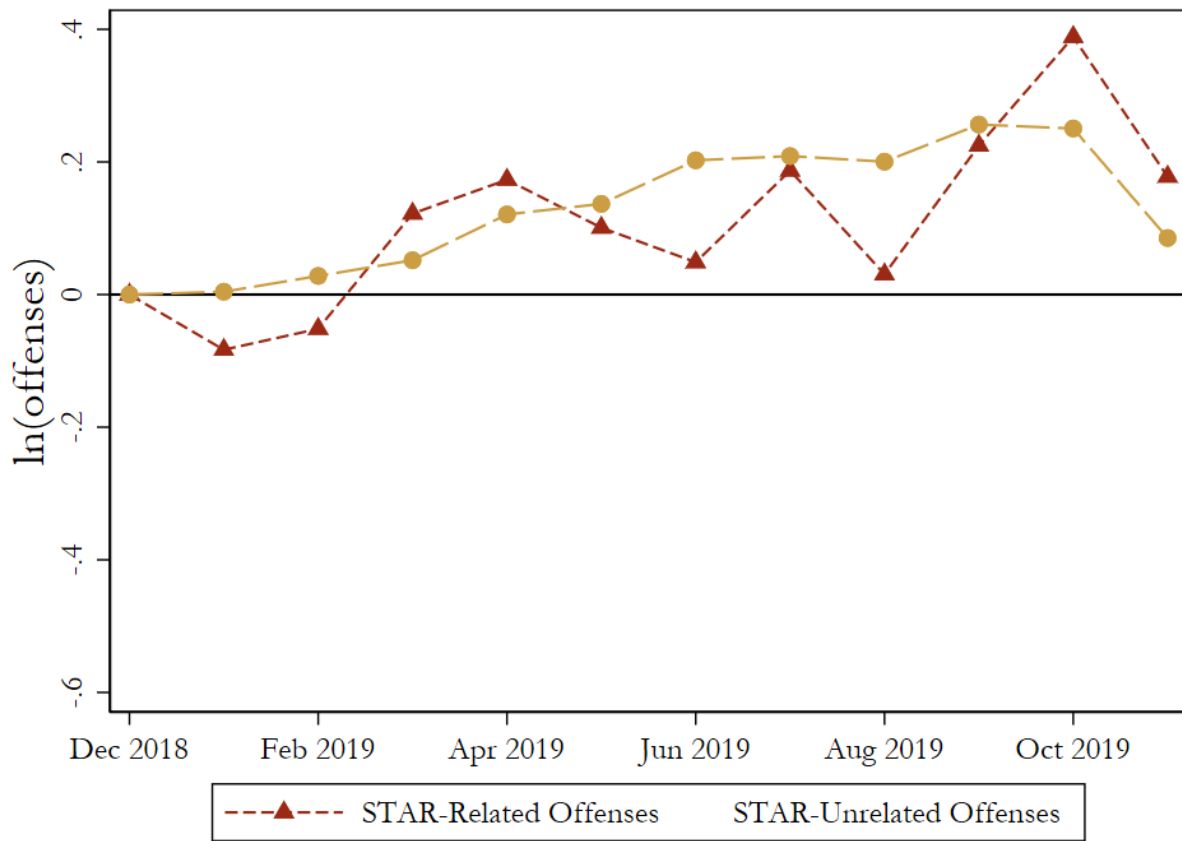


Figure S7. Placebo Event Study (Dec. 2018 – Nov. 2019). This placebo check mirrors the analyses represented in Figure 3 and Table S7, but applied to a time period other than our study window. Specifically, the difference-in-differences (DD) estimates are based on 432 precinct-month observations from December 2018 through November 2019 and condition on precinct fixed effects and month fixed effects. The outcome variables are the natural log of offense counts, differentiated by those that are STAR-related and those that are not. The horizontal line at zero denotes the baseline levels of offenses. The vertical line separates months of the year when the STAR program would and would not have been effective if the STAR program had begun in this time window rather than on June 2020.

Neighborhood	Population	% College Degree	Median HH Income	% People of Color				
				All	Black	Latinx	Asian	Other
Auraria	778	8%	\$ 86,875	30%	1%	23%	0%	6%
Baker	6,568	53%	\$ 75,973	35%	4%	24%	2%	5%
Capitol Hill	9,309	64%	\$ 63,532	14%	2%	6%	2%	4%
CBD	6,916	27%	\$ 38,888	62%	13%	44%	2%	3%
Cheesman Park	5,339	51%	\$ 60,514	32%	11%	14%	3%	4%
City Park West	4,552	26%	\$ 63,250	63%	14%	47%	0%	2%
Civic Center	18,924	46%	\$ 70,971	39%	15%	16%	2%	7%
Cole	12,061	63%	\$ 66,120	18%	2%	10%	4%	3%
Five Points	3,032	67%	\$ 197,813	8%	1%	3%	0%	3%
Lincoln Park	16,304	60%	\$ 54,762	23%	4%	11%	4%	4%
North Capitol Hill	6,754	60%	\$ 69,668	26%	6%	16%	2%	2%
Speer	2,256	59%	\$ 105,962	25%	9%	10%	5%	2%
Union Station	4,491	47%	\$ 75,323	29%	3%	16%	4%	7%
Whittier	7,500	71%	\$ 95,487	17%	1%	5%	9%	2%
All STAR	104,784	48%	\$ 72,870	33%	8%	19%	3%	4%
All Non-STAR	573,683	36%	\$ 77,596	44%	8%	30%	3%	3%
City-Wide	678,467	46%	\$ 59,179	46%	10%	33%	4%	3%

Table S1. STAR Pilot Program Neighborhood Demographics. Data come from 2019 American Community Survey data, retrieved from <https://www.denvergov.org/opendata/dataset/american-community-survey-nbrhd-2015-2019>. All the listed neighborhoods except Whittier are characterized as “displacement-vulnerable.”

STAR-Related Offenses				
Simple assault on police	Possession of cocaine	Possession of synthetic narcotic	Loitering	Property crimes - other
Simple Assault	Possession of hallucinogenic drug	False imprisonment	Obstructing government operation	Public fighting
Criminal mischief - other	Possession of heroin	Harassment	Failure to obey police order	Public order offense - other
Criminal trespassing	Possession of marijuana	Harassment of a sexual nature	Giving police false information	Public peace - other
Disarming a peace officer	Possession of methamphetamine	Indecent exposure	Police interference	Reckless endangerment
Disturbing the peace	Possession of opium or derivative	Liquor law violation - other	Obstructing criminal investigation	Threatening to injure
Possession of a barbiturate	Possession of drug paraphernalia	Possession of liquor	Resisting arrest	
STAR-Unrelated Offenses				
Accessory/conspiracy to crime	Selling cocaine	Fraud by check due to insufficient funds	Inciting a riot	Pocket picking
Aggravated assault	Forgery to obtain drugs	Gambling - betting or wagering	Robbery of a bank	Purse snatching without force
Aggravated assault - domestic violence	Fraud to obtain drugs	Running a gambling operation	Robbery of a business	Shoplifting
Altering VIN number	Manufacture of hallucinogenic drug	Gambling - gaming operation	Carjacking - armed	Theft of construction equipment
Possession of dangerous animal	Selling of hallucinogenic drug	Harassment - domestic violence	Forcible purse snatching	Theft of trailer
Arson of a business	Selling of heroine	Stalking - domestic violence	Robbery of a person in a residence	Unauthorized use of credit/debit card
Arson - other	Manufacture/sell other dangerous drugs	Conspiracy to commit homicide	Robbery of a person in the open	Habitual traffic offender
Arson to a public building	Cultivation of marijuana	Homicide by a family member	Unlawful sexual contact	Impound abandoned vehicle
Arson of a residence	Selling of marijuana	Homicide by negligence	Rape	Traffic - other
Arson of a vehicle	Selling of methamphetamine	Homicide by other means	Rape by person in position of trust	Vehicular assault
Aggravated assault to police using gun	Manufacture of methamphetamine	Homicide of a police officer with a gun	Sexual assault with an object	Vehicular homicide
Assault - domestic violence	Selling of opium or an opium derivative	Illegal dumping	Sexual assault - position of trust	Traffic accident
Bomb threat	Other dangerous drugs - PCS	Impersonation of a police officer	Sexual assault - fondling adult	Traffic accident - DUI DUID
Bribery	Selling a synthetic narcotic drug	Intimidation of a witness	Sexual assault - non rape	Traffic accident - hit and run
Burglary/auto theft of a business	Eavesdropping	Kidnap an adult	Sodomy of adult using force	Vehicular eluding
Burglary and auto theft at a business - forced entry	Escape of a prisoner	Domestic violence kidnapping	Failure to register as sex offender	Vehicular eluding - no chase
Burglary and auto theft at residence	Aiding the escape of a prisoner	Manufacture of liquor	Sex offender registration violation	Violation of court order
Burglary and auto theft at a residence - forced entry	Possession of an explosive device	Liquor - misrepresenting age	Buy, sell, receive stolen property	Violation of custody order
Burglary of a business - forced entry	Use of an explosive device	Illegal sale of liquor	Theft of bicycle	Violation of restraining order
Burglary of a business	Possession of explosive device	Littering	Theft by confidence game	Altering weapon serial number
Possession of burglary tools	Extortion	Threatening to injure with weapon	Embezzlement by an employee	Possession of weapon
Burglary of a residence - forced entry	Failure to report abuse	Money laundering	Failure to return rental vehicle	Carrying concealed weapon
Burglary of a residence	Possession of fireworks	Manufacture of obscene material	Theft from a building	Carrying prohibited weapon
Burglary of a safe	Forgery of checks	Possession of obscene material	Theft from a mailbox	Weapon fired into occupied building
Burglary of a vending machine	Counterfeiting an object	Other environmental or animal offense	Theft from a yard	Weapon fired into occupied vehicle
Smuggle contraband to a prisoner	Forgery - other	Parole violation	Theft of fuel by driving off	Flourishing a weapon
Possession of contraband	Possession of forged credit/debit card	Pawn broker violation	Theft of items from vehicle	Weapon violation - other
Criminal mischief - graffiti	Possession of a forged instrument	making a false report to police	Theft of cable services	Possession of illegal weapon
Criminal mischief to a motor vehicle	Possession of a counterfeiting device	Probation violation	Theft of motor vehicle	Unlawful discharge of weapon
Curfew violation	Fraud by telephone	Aiding the act of prostitution	Theft of rental property	Unlawful sale of weapon
Manufacture of a barbiturate	Fraud by use of computer	Engaging in prostitution	Theft of services	Window peeping
Selling a barbiturate	Criminal impersonation	Pimping for prostitution	Theft - other	Wiretapping
	Identity theft	Engaging in a riot	Theft of parts from a vehicle	Cruelty to animals

Table S2. Offense Type Codes Differentiated by STAR-Related and STAR-Unrelated offenses. Offenses were differentiated using two independent coders who interpreted STAR guidelines for dispatching clinicians to mental-health and substance-abuse calls.

Variable	Mean	SD	Minimum	Maximum
Active STAR Program	0.11	0.31	0	1
STAR-related offenses	33.7	30.9	5	258
STAR-unrelated offenses	159.6	54.6	27	406
Total offenses	193.3	75.3	48	606
<i>Offenses by category</i>				
Alcohol and drugs	6.3	7.6	0	64
Assault	6.8	4.9	0	26
Auto theft	18.7	10.5	1	66
Burglary	11.8	6.8	0	42
Disorderly conduct	22.6	14.6	2	168
Larceny	22.7	15.2	1	87
Theft from motor vehicle	23.2	12.6	1	88
Traffic offenses	35.7	19.0	1	119
Other crimes against people	9.3	6.7	0	43
All other offenses	28.8	24.9	1	204

Table S3. Descriptive Statistics. The sample is based on Denver's 36 police precincts observed in each of 12 months from December 2019 through November 2020 (n = 432 precinct-month observations). Due to the very low instances of arson (0.2%), murder (0.1%), robbery (2.8%), sexual assault (1.5%), and white-collar offenses (2.7%), we do not include those STAR-unrelated offense categories in the table.

Independent Variables	DD				DDD	
	STAR-Related Offenses		STAR-Unrelated Offenses			
	(1)	(2)	(3)	(4)	(5)	(6)
Active STAR program	-0.41*** (0.07)		-0.05 (0.04)		-0.36*** (0.05)	
Adoption month: June 2020		-0.34*** (0.12)		0.01 (0.06)		-0.34*** (0.09)
1-month lag: July 2020		-0.36*** (0.11)		-0.02 (0.06)		-0.33*** (0.10)
2-month lag: August 2020		-0.39*** (0.10)		-0.06* (0.04)		-0.32*** (0.09)
3-month lag: September 2020		-0.47*** (0.11)		-0.04 (0.05)		-0.43*** (0.10)
4-month lag: October 2020		-0.45*** (0.08)		-0.07 (0.06)		-0.38*** (0.10)
5-month lag: November 2020		-0.48*** (0.11)		-0.10 (0.06)		-0.38*** (0.10)
<i>p</i> value ($H_0: \delta_0 = \delta_{-1} = \delta_{-2} = \delta_{-3} = \delta_{-4} = \delta_{-5}$)		0.91		0.68		0.90

Table S4. Estimated Effects of the STAR Program on Criminal Incidents. The difference-in-differences (DD) estimates are based on 432 precinct-month observations and condition on precinct fixed effects and month fixed effects. The difference-in-difference-in-differences (DDD) estimates are based on the stacked precinct-month data for STAR and non-STAR offenses ($n = 864$). The DDD estimates condition on fixed effects unique to each category of the following 2-way interactions: precinct-by-month, precinct-by-STAR offense, and month-by-STAR offense. The outcome variables are the natural log of the offense counts. Standard errors, clustered at the precinct level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Outcome Variable	DD Estimate
All offenses	-0.15*** (0.05)
Alcohol and drugs	-0.53** (0.20)
Assault	0.05 (0.15)
Auto theft	0.15* (0.09)
Burglary	-0.06 (0.09)
Disorderly conduct	-0.20*** (0.06)
Larceny	-0.06 (0.08)
Theft from motor vehicle	0.04 (0.09)
Traffic offenses	-0.04 (0.06)
Other crimes against people	-0.14* (0.07)
All other offenses	-0.13 (0.12)

Table S5. Estimated Effects by Offense Category. The difference-in-differences (DD) estimates are based on 432 precinct-month observations and condition on precinct fixed effects and month fixed effects. Due to the very low instances of arson, murder, robbery, sexual assault, and white-collar offenses, we do not include those STAR-unrelated offense categories. The outcome variables are the natural log of the offense counts. Standard errors, clustered at the precinct level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Model	DD Estimate
Clustering corrected for finite sample bias	-0.41*** (0.07)
Poisson fixed effects model	-0.45*** (0.03)
Negative binomial fixed effects model	-0.42*** (0.06)
Without Precinct 311 (n=420)	-0.43*** (0.07)
Post-pandemic months (March-Nov 2020) only (n=324)	-0.38*** (0.06)
Effects during STAR-eligible days and times	-0.37*** (0.13)
Effects during STAR-ineligible days and times	-0.44*** (0.06)
Recoding simple assaults as STAR-unrelated offenses	-0.49*** (0.08)
Recoding May 2020 as a treatment month	-0.36*** (0.09)
Effects in STAR-adjacent precincts	0.02 (0.08)
Placebo effects, June - Nov. 2017 (Dec. 2016 start)	0.08 (0.08)
Placebo effects, June - Nov. 2018 (Dec. 2017 start)	0.04 (0.09)
Placebo effects, June - Nov. 2019 (Dec. 2018 start)	0.13* (0.05)

Table S6. Alternative specifications. The difference-in-differences (DD) estimates are based on 432 precinct-month observations (unless otherwise noted). All models condition on month fixed effects. All models condition on precinct fixed effects. The outcome variables are the natural log of the STAR-relevant offense counts. Standard errors, clustered at the precinct level, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Independent Variables	Offense Type	
	STAR- Related	STAR- Unrelated
	(1)	(2)
5-month lead	0.02 (0.12)	0.03 (0.05)
4-month lead	-0.02 (0.10)	0.01 (0.05)
3-month lead	-0.11 (0.12)	0.03 (0.07)
2-month lead	-0.15 (0.13)	0.03 (0.12)
1-month lead	-0.07 (0.24)	0.12 (0.13)
Precinct participated in June 2020	-0.39** (0.19)	0.04 (0.10)
Precinct participated in July 2020	-0.41** (0.17)	0.01 (0.09)
Precinct participated in August 2020	-0.44*** (0.16)	-0.03 (0.07)
Precinct participated in September 2020	-0.52*** (0.17)	-0.00 (0.08)
Precinct participated in October 2020	-0.50*** (0.13)	-0.04 (0.09)
Precinct participated in November 2020	-0.53*** (0.14)	-0.06 (0.08)
p value ($H_0: \delta_5=\delta_4=\delta_3=\delta_2=\delta_1=0$)	0.71	0.87
p value ($H_0: \delta_3=\delta_2=\delta_1=0$)	0.60	0.62
p value ($H_0: \delta_0=\delta_{-1}=\delta_{-2}=\delta_{-3}=\delta_{-4}=\delta_{-5}$)	0.91	0.69

Table S7. Event-study estimates. The difference-in-differences (DD) estimates are based on 432 precinct-month observations and condition on precinct fixed effects and month fixed effects. The outcome variables are the natural log of the STAR-related offense counts. Standard errors, clustered at the precinct level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

	GSC		CITS	
	STAR Offenses	Non- STAR Offenses	STAR Offenses	Non- STAR Offenses
Treatment X Post	-0.46*** (0.14)	0.02 (0.11)	-0.24** (0.10)	-0.03 (0.05)
Treatment X Post X Trend	--	--	0.00 (0.06)	-0.04 (0.03)
Treatment X Trend	--	--	-0.03 (0.04)	0.02 (0.03)

Table S8. Comparative interrupted time series estimates. The dependent variable is the natural log of the stated offenses ($n = 432$ precinct-month observations). The first two columns report estimates based on generalized synthetic control (GSC; Xu, 2017, Liu et al., 2020) and bootstrapped standard errors (1,000 replications). The next two columns report estimates based on a comparative interrupted time-series (CITS) specification (Shadish, Cook, and Campbell, 2002) and standard errors clustered at the precinct level. The CITS specifications also condition on precinct fixed effects and month fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.