HOUSING AND COMMUNITIES

Aligning Crisis Response with Community Needs

Evidence from Denver's Support Team Assisted Response (STAR) and Co-Responder Programs

Sarah Gillespie **Will Curran-Groome** **Brendan Chen** **Devlin Hanson**

URBAN INSTITUTE

RESEARCH REPORT

February 2026

ABOUT THE URBAN INSTITUTE

The Urban Institute is a nonprofit research organization founded on one simple idea: To improve lives and strengthen communities, we need practices and policies that work. For more than 50 years, that has been our charge. By equipping changemakers with evidence and solutions, together we can create a future where every person and community has the opportunity and power to thrive.

Contents

Acknowledgments

This report was funded by the City and County of Denver with support from the Caring for Denver Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

This study benefited greatly from the substantial contributions of time and technical expertise from the many City of Denver staff and partners who lead the STAR and Co-Responder programs. The authors also gratefully acknowledge Mari McGilton Bayer and Amy Rogin, former Urban staff, for their critical research support during earlier phases of this study.

Executive Summary

Cities across the United States are developing community responder models—where trained mental health practitioners and behavioral health clinicians respond to noncriminal emergencies in lieu of or in addition to police units—to better serve people experiencing mental health issues, substance use challenges, homelessness, and other crises. Such models aim to improve outcomes for individuals in crisis, reduce unnecessary law enforcement involvement, and better align emergency response systems with public health and community-based services. Denver has invested substantially in two approaches: the Support Team Assisted Response (STAR) program and the Co-Responder Program. These programs offer distinct but complementary pathways for responding to behavioral health–related 9-1-1 calls**.**

This report presents a comprehensive evaluation of Denver's STAR and Co-Responder programs. Using program data linked with administrative records from the Denver Department of Public Safety, we examine program implementation, scope and scale, individual-level outcomes in the year following a STAR or co-responder encounter, implications for cost and scalability, and priorities for future research and program improvement. The study builds on our prior qualitative work describing each of the two programs and, to our knowledge, provides the first estimates of individual-level outcomes during the year following an alternative response or co-responder response.

Program Models and Implementation

STAR and the Co-Responder Program both seek to improve responses to behavioral health–related 9-1-1 calls by providing on-scene assessment, de-escalation, and other forms of support. However, they differ in structure and operational roles.

STAR is a civilian-led crisis response model in which unarmed clinical van teams—paramedics or EMTs and behavioral health professionals—respond in lieu of police to eligible, low-risk calls. Following these on-scene interactions, STAR responders may refer clients to the STAR Community Partner Network, which provides culturally, linguistically, and geographically responsive case management services and follow- up care. STAR launched as a pilot in downtown Denver in 2020 and expanded citywide in 2022. By 2024, Denver 9-1-1 designated approximately 1,000 to 1,500 calls for service per month as STAR-eligible, out of 45,000 total monthly 9-1-1 calls for service for police, fire, and medical services. Over the four and a half years included in this study—from June 2020 through September 2024—Denver 9-1-1 designated almost 50,000 calls for service as STAR-eligible. Due to limited

capacity, STAR van teams could respond to only a portion of those calls. Police responded to 57 percent of STAR-eligible calls either alone or with another public safety agency, while STAR vans responded to 21 percent without police involvement and an additional 6 percent of calls that police also responded to. In total, STAR vans responded to 23,114 calls for service from 9-1-1.

The Co-Responder Program is a police–clinician partnership model in which mental health clinicians respond alongside police officers, often in situations where police involvement remains necessary. Launched in 2016 and expanded citywide by 2018, the program operates 24/7 (for the period covered in this report: June 2020 – September 2024)[1] and addresses a broad range of behavioral health–related calls. Clinicians support on-scene assessment, de-escalation, and service connection and may continue engagement with clients after the initial encounter. During the study period, co-responders accompanied police on an average of 7,664 responses per year.

Across both programs, implementation data show that responses are concentrated in certain districts—particularly downtown and north-central Denver—and frequently involve people with prior law enforcement contact and experiences of homelessness. At the same time, individuals identifying as Hispanic or Latino/a/x are significantly underrepresented among STAR and Co-Responder clients when compared with the city population, suggesting persistent gaps in access and unmet need, although the trends require further analysis to understand the root causes. Together, these implementation findings highlight both the scale of Denver's alternative and co-responder efforts and the operational constraints that shape who is reached, how often teams respond, and where additional capacity or targeted outreach could further strengthen program impact.

Impacts on Criminal Justice Outcomes

A central contribution of this study is its examination of individual-level outcomes in the year following STAR and co-responder encounters. We describe how these two community responder programs affect individuals' post-encounter criminal-legal contacts, including post-encounter police contacts, arrests, and bookings.

Methods and Limitations

Our goal was to quantify how receipt of services from STAR or Co-Responder units in response to a 9-1-1 call for service (the treatment group) affected individuals' outcomes as compared with the status quo, where a police unit would respond to the call for service (the comparison group). Ideally, we could

simply compare outcomes for individuals who received treatment against those who had identical types of calls-for-service but who due to capacity limitations received a status quo police response.

However, 9-1-1 call data do not include individual identifiers, which we need to link calls back to Denver Department of Public Safety (DOS) records for police contacts, arrests, and bookings; these DOS records are critical because we use them to define our outcomes. Instead, we identified the treatment group based on clinical encounters documented in program data, and a comparison group directly from DOS records. Our approach is premised on the idea that a 9-1-1 call for service that does not receive a response from a STAR or Co-Responder unit will instead receive a police response reflected in DOS records. DOS records encompass police responses for calls-for-service that are similar to the calls-for-service to which STAR and Co-Responder units respond—for example, for events like welfare checks and disturbances—and calls-for-service that STAR and Co-Responder units might not respond to, such as situations involving violence. However, we needed to identify only comparable DOS records to create our comparison group. We did this using an approach called propensity score matching, using characteristics such as race, age, prior experiences of homelessness, the reason for the encounter, and past histories of criminal justice interactions to create a comparison group that was as similar as possible to our treatment groups.

With our treatment groups and our matched comparison groups, we then ran a series of regressions to examine what effects treatment had on individuals' criminal-justice encounters in the following year. We quantified effects on police contacts, arrests, bookings, and subsequent STAR and co-responder encounters, and we looked at these effects for subgroups, such as for all individuals with prior experiences of homelessness.

This approach included a number of important limitations. We could only measure STAR and co-responder encounters that were documented as clinical encounters. Clinical encounters are meant to reflect any substantive engagement with a client, but data collection challenges limited the number of these encounters documented by program staff. We had limited data available on which to "match" our DOS comparison group records to our treatment group records. This means that some of the encounters in our comparison group were likely not eligible for a STAR response, though we cannot tell which ones or how many. We also did not have robust data about the types of post-encounter case management and other supportive services that STAR and co-responder clients may have received, despite the fact that these services are a critical part of the program models. Lastly, we only had data to measure outcomes related to criminal justice encounters; we were unable to evaluate other important outcomes of interest, such as health, housing, and employment outcomes.

STAR Outcomes

The study finds statistically significant reductions in subsequent criminal justice involvement during the year following a first-time STAR van clinical encounter. Compared with people who received a police response for similar types of encounters, STAR clients were less likely to experience any police contacts and any arrests in the year after their STAR encounter. Although changes in aggregate counts of arrests and bookings were modest at the individual level, reductions in police contacts were meaningful when considered across the volume of STAR encounters delivered annually. Clients were also substantially more likely to experience subsequent STAR encounters, suggesting increased connection to services rather than reverting back to a pattern of traditional police responses.

We examined whether outcomes were different for people with different characteristics. The impacts were amplified for people with higher average levels of system involvement—most notably people who had experienced homelessness prior to their encounter. For this group, reductions in police contacts and arrests were two to three times larger than those observed for the full STAR population, translating into substantial aggregate reductions in criminal justice interactions at the program level.

Co-Responder Outcomes

The Co-Responder Program also demonstrates significant impacts on criminal justice outcomes. Individuals who experienced a clinical co-responder encounter were significantly less likely to be arrested during the year following the encounter than comparable individuals who experienced a police-only response. Reductions were also observed in police contacts and bookings, and clients were substantially more likely to receive subsequent clinical encounters—suggesting increased connection to services rather than reverting back to a pattern of traditional police responses.

As with STAR, the largest effects were observed among people who had experienced homelessness. Although this subgroup represents a minority of overall co-responder clients, their higher average arrest and jail rates mean that reductions in these rates produce far larger absolute changes than in the overall co-responder client population. When aggregated across the program's annual volume of encounters, those reductions represent meaningful decreases in police contacts, arrests, and bookings, alongside associated cost offsets.

STAR Cost and Scalability Analyses

The report also examines STAR program costs, potential cost offsets, and implications for scalability. In 2023, annual STAR program costs—including van teams and follow-up services—were approximately $4.4 million, including $532,000 in capital costs for new vans. This translates to an estimated $237 per response for van teams alone and $470 per response when accounting for all program components including the Community Partner Network. We also discuss important in-kind resources required for STAR operations. While STAR is more resource- intensive than a police contact, it is designed to provide clinical assessment, de-escalation, and service connection—services that police contacts alone are not structured to deliver. Furthermore, the analysis shows that even modest reductions in arrests and subsequent bookings can generate meaningful cost avoidance. Looking ahead, the report finds that if the program could operate nine to ten concurrent STAR van units per day, at the current average of about five responses per unit per day, the program could serve the current volume of need for STAR services. This estimate of need includes both 9-1-1 calls flagged as STAR-eligible and the STAR responses requested directly by other agencies.

Key Implications and Recommendations

Taken together, the findings underscore the value of maintaining multiple crisis response pathways rather than relying on a single model. STAR and the Co-Responder Program serve overlapping populations through different mechanisms: STAR expands the city's capacity to respond without police involvement, while the Co-Responder Program improves outcomes in situations where police presence remains necessary. Their combined impact is strongest for individuals with high levels of system involvement, reinforcing the importance of aligning the best response to each crisis based on need.

The report identifies several opportunities to strengthen both programs, including deepening understanding of access for underrepresented communities, increasing the number of encounters and documentation of encounters, expanding capacity to operate more concurrent van units each day to meet demand, exploring opportunities to increase responses per day per STAR van team, and investing in data systems that track referrals, service engagement, and longer-term outcomes. Because service connection is central to both programs' theories of change, addressing current data limitations to assessing service engagement and other health and well-being outcomes is essential to understanding their full impact.

Conclusion

This evaluation provides compelling evidence that Denver's STAR and Co-Responder programs reduce subsequent criminal justice involvement, particularly for people with high system contact, while offering service-oriented responses to behavioral health crises. By combining rigorous individual-level outcome analysis with detailed implementation, cost, and scalability assessments, the report advances the national evidence base on community responder models and offers actionable insights for program leaders and policymakers. The findings demonstrate that both community responder models can meaningfully reduce reliance on the criminal-legal system while better matching individuals in crisis to appropriate services. The study underscores the importance of a diversified crisis response system that aligns service intensity with client need and supports continued learning and improvement.

Aligning Crisis Response with Community Needs

Across the United States, cities are developing varying community responder models to better address 9-1-1 calls involving mental health crises, substance use, and related needs that may not require a traditional law enforcement response. In Denver, two distinct but complementary approaches—the Support Team Assisted Response (STAR) program and the Co-Responder Program—anchor the city's crisis responder efforts. STAR is a civilian-led model in which clinical van teams respond in lieu of police for eligible calls and make referrals for case management and follow up services that are provided through a Community Partner Network; the Co-Responder Program pairs clinicians with police officers to provide on-scene clinical expertise during encounters where police involvement remains necessary. Together, these programs aim to reduce reliance on arrest and detention, improve outcomes for people in crisis, and better align emergency response with behavioral health and community-based services.

This paper evaluates the implementation and impacts of Denver's STAR and Co-Responder programs using program data and linked administrative records from the Denver Department of Public Safety (DOS). We describe the scope, scale, and operational characteristics of each model and estimate their effects on subsequent criminal justice outcomes for individuals in the year following a STAR or co-responder encounter. We also examine differential impacts for key populations, discuss implications for scalability and cost, and identify limitations and priorities for future research. Taken together, the findings contribute to the growing evidence base on alternative crisis response models and offer practical insights for policymakers and program leaders seeking to strengthen community-based crisis response systems.

Methods

To estimate the impacts of each of the STAR and Co-Responder programs on criminal-legal system engagement, we compared criminal-legal encounters following a person's first clinical encounter with either the STAR or Co-Responder program to the criminal-legal encounters of similar people who instead had a police contact. We could only measure STAR and Co-Responder encounters that were documented as clinical encounters. Clinical encounters are meant to reflect any substantive engagement with a client, but data collection challenges limited the number of these encounters documented by program staff. We looked at criminal-justice encounters including police contacts, arrests, bookings, and STAR and Co-Responder clinical encounters during the year following each focal

encounter (either the clinical encounter or a similar police contact). It is important to note that we do not look at any same-day outcomes—any police response that was avoided on the day of the encounter because STAR responded is not counted as part of the outcome.

Ideally we would have identified both the clinical encounters and the police 9-1-1 encounters from the Computer-Aided Dispatch (CAD) data; receipt of a clinical response would be pseudo-random as a function of whether a STAR van or Co-Responder unit was available when a call for service was received, and the treatment group (STAR or co-responder clinical encounters) would look virtually identical to the comparison group (encounters that could have received a clinical encounter but received a police response because STAR vans or Co-Responder units were not available).

However, because the CAD data do not include personal identifiers, we were unable to link CAD data to individual criminal-justice encounters. Instead, we used program data from WellPower—the organization that provides clinical staff for STAR and the Co-Responder Program--about clinical encounter clients, which the City of Denver was able to link to DOS data. To best identify a comparison group of police encounters that was similar to the treatment group, we leveraged public safety data. Our approach is premised on the fact that if someone called 9-1-1 but did not receive a STAR or co-responder response, the police would likely have responded with a field interview, the lowest level of police contact used when there is no likelihood of violence, crime, or other serious concern. (Because Co-Responder units can respond to any call for service, not just the more limited set of "STAR-eligible" calls for service that STAR van teams respond to, we defined the co-responder comparison group slightly differently. Refer to Appendix A for additional details.)

For the evaluation of STAR outcomes, we further narrowed the comparison group to field interviews with reason codes used in the public safety data that were similar in nature to the STAR-eligible codes used in the CAD data. This approach was limited by the imperfect alignment across CAD and public safety data, and as a result, some of our comparison encounters may not have been responses to 9-1-1 calls or STAR-eligible. In addition, we may have excluded other encounters from the comparison group that were in fact STAR-eligible. The comparison group reflects the best alignment we could produce using the public safety data. Like the treatment group, the comparison group also only includes police encounters for people who had never before had a STAR clinical encounter.

After identifying the clinical encounters and police encounters to be used in our analyses, we used propensity score matching to identify or "match" treatment and comparison group encounters that were as similar as possible in terms of their encounter year, encounter reason, client age, client race, client homelessness history, and client history of prior police encounters including arrests, general

occurrences, field interviews, and bookings. Using these methods, the differences between the outcomes experienced by people represented in the two groups (clinical encounters and police encounters) in the year following each encounter can be attributed to the impact of the STAR or co-responder clinical encounter. (See Appendix A, figure A1 for an illustration of this analysis method and box 1 for more information on interpreting statistical results.)

BOX 1

Interpreting Statistical Results

We use regression models to evaluate the impacts of both STAR and co-responder clinical encounters on individuals' criminal justice outcomes. For each model, we present a number of statistics to help readers understand different dimensions of the relationships in question, including measures of statistical significance and effect size.

Statistical significance tells us the likelihood the difference observed in a regression is due to random chance or the impact of the intervention. However, statistical significance is partly a reflection of sample size. Results can be statistically significant even when the difference is very small when there are many observations.

Effect sizes tell us about whether the magnitude of a result is meaningful. Consider a statistically significant result based on an analysis of 1 million observations. The magnitude of the relationship in question could be very small but still be statistically significant because we're looking at so many observations. An effect size contextualizes that magnitude by comparing it to the distribution of the outcome variable.

For example: we run a regression to evaluate whether a STAR clinical encounter—compared with a field interview with police—influences the number of arrests a person experiences in the year following their encounter. Let's imagine that arrests are a fairly rare event—only 1 in 10 people in our population of interest has an arrest in a given year. If we then find that a STAR clinical encounter reduces the number of arrests in the following year by an average of 0.05 arrests per person, this would yield a fairly large effect size—0.05 arrests might not sound like a lot, but the average person only has 0.1 arrests per year. By contrast, if arrests are a fairly common event, with each person in our population having 3 arrests per year on average, a reduction of 0.05 arrests per person per year suddenly sounds much less impactful. In both scenarios we might find statistical significance because we have reasonably large numbers of observations, but in the latter scenario, we would likely have a very small effect size, suggesting the results aren't very impactful.

We look at the effects of receiving a clinical encounter in lieu of a police field interview on individuals' subsequent events—police contacts, arrests, bookings, and other STAR and co-responder

clinical encounters—in the year following their first clinical encounter with each program. We explore the effects on continuous (count) outcomes, such as the number of arrests in the following year, as well as binary outcomes, such as whether an individual had *any* arrest in the following year. These approaches provide complementary pictures of the effects of the programs, with binary outcomes describing the share of people who, for example, avoided arrest in the following year, and with continuous outcomes describing the average number of arrests avoided.

Support Team Assisted Response (STAR) Findings

Alternative response programs have expanded nationwide as a promising approach to handling emergency calls related to mental health crises, substance use, homelessness, and other noncriminal concerns. Programs such as Denver's STAR, Albuquerque's Behavioral Health Responders, San Francisco's Street Crisis Response Team, Durham's HEART, St. Petersburg's CALL, and Portland Street Response typically deploy unarmed teams—clinicians, paramedics, social workers, and peers—to provide de-escalation, crisis stabilization, and service connections instead of police intervention. Case studies consistently show such teams can resolve many behavioral health calls on scene, often more quickly than police, with extremely low rates of arrest, use of force, or escalation (Townley and Leickly 2021; City and County of San Francisco Street Response Team 2022).[2] Qualitative research highlights strong community support, particularly among individuals with lived experience of homelessness or behavioral health challenges, who report feeling safer and more respected when alternative responders arrive (Gillooly et al. 2023; Townley et al. 2019).

Despite these successes, implementation challenges persist. Police often remain the default responders due to limitations in call-taking and dispatch systems, restricted program hours or staffing for alternative response programs, and organizational or cultural barriers to referrals (Gillooly et al. 2023). Effective scaling requires more than deploying new teams—it depends on integrating programs into 9-1-1 systems, standardizing data collection, fostering cross-agency partnerships, and securing sustainable funding. Although rigorous evaluations remain limited, recent evidence suggests that well-integrated, adequately resourced, and community-informed models can build trust, reduce police burden, strengthen service connections, and better align emergency response with public health goals (Yamamoto et al. 2025; Gonzalez Miranda et al. 2024).

To date, most research has relied on qualitative and descriptive methods, focusing on who responds, how often, and what occurs during the call (Yamamoto et al. 2025; Gonzalez Miranda et al. 2024; Beck et al. 2020). Few studies examine follow-up services, and even fewer measure service uptake or retention. One 2022 study estimated community-level crime impacts in Denver, and a recent 2025 evaluation of Oregon's CAHOOTS (Crisis Assistance Helping Out on the Streets) program is the first to estimate causal outcomes at the call level (Dee and Pyne 2022; Davis et al. 2025). That study also analyzed repeat 9-1-1 calls, incorporated cost-effectiveness, and explored optimal service expansion. However, no research has yet measured individual-level outcomes post-response. Data limitations—particularly with computer-aided dispatch (CAD) systems designed for operations rather than research—pose major challenges. CAD data are designed to address questions of "where?" and "when?" so that first responders can be effectively dispatched to address crises; these data often lack identifiers for linking individuals across systems, making it complicated to answer research questions focused on individual outcomes ("who?" and "what?"). Research gaps also reflect outcomes that are more challenging to measure with administrative data. Unlike measuring arrests and jail stays, measuring service connections and well-being is difficult without primary data sources. Other important dimensions of alternative response programs include their cost-effectiveness and their ability to scale. However, identifying appropriate measurements of these issues can be challenging and risks misinterpreting the role of alternative response within an ecosystem of first responders. For example, focusing on quantifying the costs associated with avoided criminal-legal encounters may miss important health and well-being benefits, resulting in underestimations of the value of alternative response programs. This is one important limitation of this study: Due to insufficient data availability, we cannot measure outcomes beyond criminal-legal system encounters.

We began studying Denver's alternative response program—STAR—in 2022. Our published work includes an implementation study and descriptive interim findings (Gillespie et al. 2023b; Gillespie et al. 2024). In this paper, we present results describing STAR implementation to date—from June 2020 through September 2024—and individual-level one-year outcomes, a cost analysis, an analysis of program costs, and scalability considerations. To our knowledge, this is the first study to examine individual-level outcomes following alternative response.

Implementation

Denver has been operating STAR since 2020. The program has multiple components. STAR pairs paramedics or EMTs and behavioral health professionals on designated vans to respond to certain 9-1-1

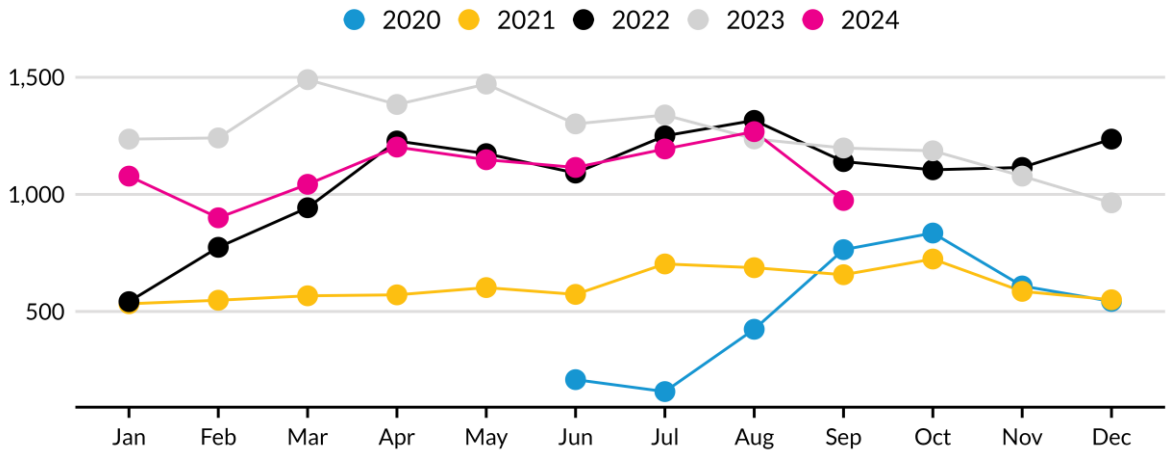
calls in lieu of a conventional response from law enforcement or other first responders. After responding to the immediate crisis, van teams make referrals to the STAR Community Partner Network for case management services and follow-up care based on client needs. The STAR Program offers an alternative to having police respond to non-crime calls such as behavioral health crises—reducing the potential for clients to experience additional trauma and criminal justice interactions—and connects clients with resources that can help them find stability and address their longer-term service needs. This paper focuses primarily on the crisis responder component of the STAR program. It has been in operation the longest and can be studied using existing individual-level program data and other administrative data.

STAR started in 2020 as a pilot program in downtown Denver and expanded citywide in early 2021. STAR is dispatched through Denver's 9-1-1 system to calls with low risk for violence or harm to STAR teams, the client, or the public through a protocol that designates calls for service that are STAR-eligible. By 2024, Denver designated around 1,000 to 1,500 calls for service each month as STAR-eligible (figure 1). For comparison, Denver 9-1-1 generates about 45,000 calls for service each month for police, fire, and medical services.[3]

FIGURE 1

Monthly STAR-Eligible Calls for Service Have Held Fairly Steady Since Full Implementation in Early 2022

STAR-eligible calls for service by year and month (N = 48,804)



URBAN INSTITUTE

Source: Authors' analyses of STAR-eligible call-for-service data provided by the City and County of Denver.
Note: Data for this figure include STAR services from June 2020 (beginning of STAR program implementation) to September 2024.

Over the four and a half years included in this study, Denver 9-1-1 designated almost 50,000 calls for service as STAR-eligible. Due to limited capacity, STAR van teams could not respond to all of these calls. In total, STAR vans responded to 23,114 total calls for service.[4] A little more than half of those calls (57 percent) originated as STAR-eligible calls for service to 9-1-1; the other 43 percent originated as calls for service that were not initially flagged as STAR-eligible but where STAR was later identified as the right resource for the call by another agency. Included in these STAR-eligible calls is a small share (about 8 percent) where the caller specifically requested a STAR response, a type of call that Denver started tracking in 2023. STAR responded to two-thirds of such calls where STAR was specifically requested.

Some calls required responses by more than one public safety agency, which could include any combination of police, fire, emergency medical service (EMS,), or STAR, and these agencies could be on scene at the same time or at different times during the response. Police responded to 57 percent of STAR-eligible calls alone or with another public safety agency that wasn't STAR. STAR vans responded to 21 percent alone or with another public safety agency that wasn't police, and to an additional 6 percent of calls that police also responded to. Other public safety agencies responded to 2 percent of these calls, and the CAD data did not document a responding agency for the remaining 14 percent of calls (table 1).

Examining the 6 percent of all STAR-eligible calls where STAR and police or other first responders were on scene at some point highlights how STAR provides a resource that all public safety agencies utilize. Of those responses with listed arrival times, police were first on the scene and requested STAR for additional resources and support for 54 percent of these calls. That could happen because the call was not initially marked as eligible for a STAR response but later was identified as appropriate for a STAR response or because there was a need for STAR-specific support and resources. For 17 percent of these calls, STAR arrived after police cleared the scene. Another share of these calls had police arrive after STAR was already on scene or had cleared the scene (28 percent) or had STAR and police arrive at the same time (1 percent). Police can arrive on scene after STAR because clients can request to speak with a police officer, including to make a formal police report, as one example.

TABLE 1

In 2024, STAR Van Teams Alone Responded to More Than One in Five STAR-Eligible Calls for Service—The Highest Share of Calls for Any Year to Date

STAR-eligible calls for service by responding agency by year (N = 48,804)

Responding agencies	2020	2021	2022	2023	2024*	Total
STAR alone	391 (11%)	559 (8%)	2,233 (17%)	3,695 (24%)	2,923 (29%)	9,801 (20%)
STAR and police	67 (2%)	176 (2%)	941 (7%)	1,032 (7%)	551 (6%)	2,767 (6%)
STAR then police	8	39	146	181	111	485
Same time	1	1	2	6	3	13
Police then STAR	22	56	382	327	131	918
Police cleared, then STAR	2	8	91	90	96	287
Missing arrival time	34	72	320	428	210	1,064
STAR and nonpolice agency	14 (0%)	20 (0%)	146 (1%)	255 (2%)	194 (2%)	629 (1%)
Police, alone or with another non-STAR agency	2,414 (68%)	5,121 (70%)	7,752 (60%)	7,747 (51%)	4,707 (47%)	27,741 (57%)
Other agency	38 (1%)	112 (2%)	275 (2%)	272 (2%)	142 (1%)	839 (2%)
No vehicles arrived	618 (17%)	1,313 (18%)	1,567 (12%)	2,124 (14%)	1,405 (14%)	7,027 (14%)

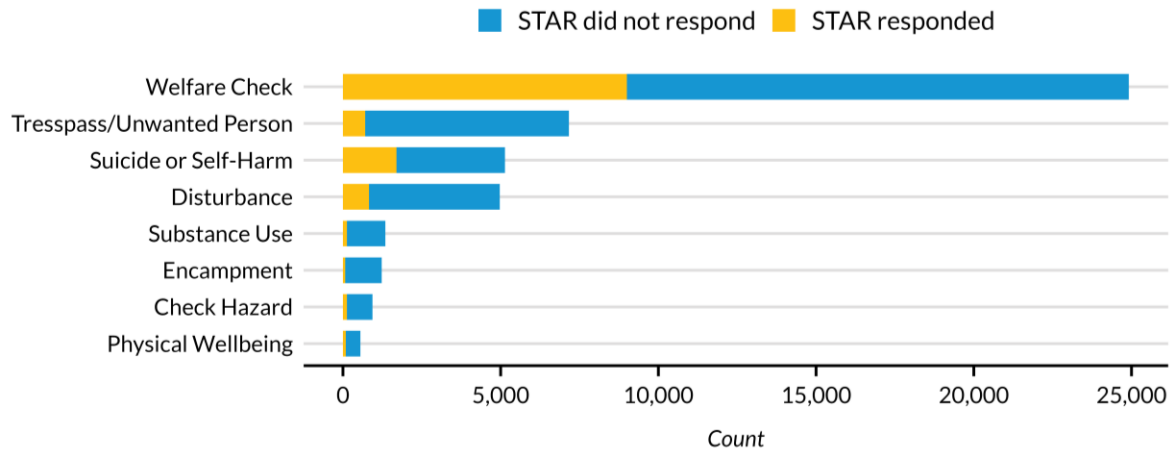
Source: Authors' analyses of STAR-eligible call-for-service data provided by the City and County of Denver.
Notes: Data include STAR services from June 2020 (beginning of STAR program implementation) to September 2024. "Non-STAR agency" includes the fire department, emergency medical service (referred to as "EMS" in the data), detoxification (referred to as "DTX" in the data), and animal protection (referred to as "DAP" in the data).
*Data from 2024 are partial-year data; counts are not directly comparable to counts from other years where data reflect responses for a full year.

Among all STAR-eligible calls, the most common nature code was recorded as a welfare check, and those were also the majority of STAR responses (figure 2).

FIGURE 2

Most STAR-Eligible Calls Are for Welfare Checks

STAR-eligible calls for service by nature code and STAR response (N = 48,804)



URBAN INSTITUTE

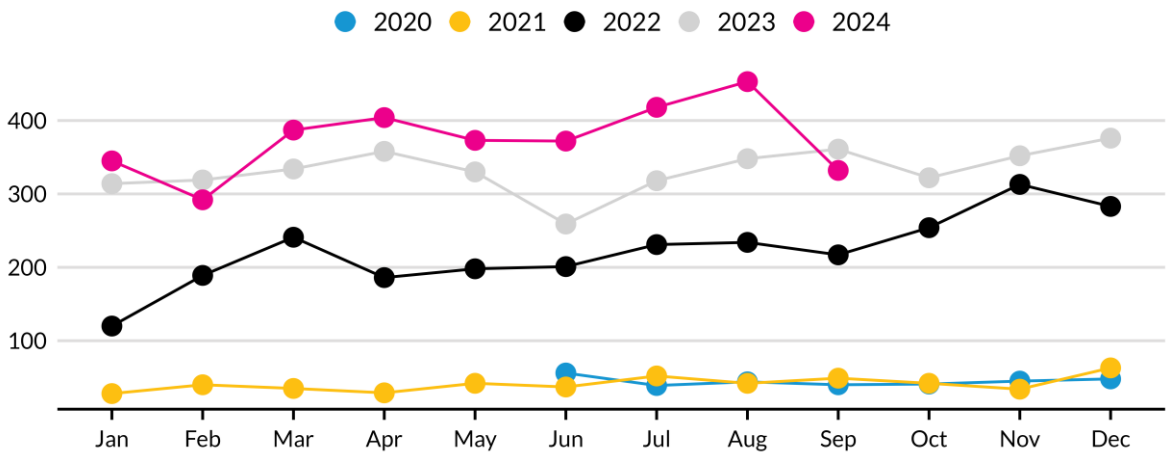
Source: Denver Department of Public Safety.
Note: Data include STAR services from June 2020 (beginning of STAR program implementation) to September 2024.

About half of all STAR van responses resulted in a substantive engagement with the STAR van teams, which the program recorded as clinical encounters. Unlike 9-1-1 calls for service and the full universe of STAR responses, STAR van teams collect identified data about clinical encounters. While data collection challenges may have limited the number of documented clinical encounters during the four-and-a-half-year study period, data showed 10,840 clinical STAR encounters, which represent almost half of all STAR responses (figure 3). Those clinical encounters served 6,664 unique clients. Approximately one in four people with a clinical STAR encounter had multiple clinical STAR encounters, and a small subset of high-frequency clients (66 clients) had 10 or more clinical STAR encounters. At the highest end, one client had 63 clinical STAR encounters during this period, an average of more than one encounter a month.

FIGURE 3

STAR Has Responded to Hundreds of Calls Each Month Since Full Implementation in Early 2022

Clinical STAR encounters by month and year (N = 10,840)



URBAN INSTITUTE

Source: Authors' analyses of STAR encounter data from WellPower.
Notes: Data include STAR services from June 2020 (beginning of STAR program implementation) to September 2024. Accordingly, counts of STAR-eligible calls for service were not available for January to May 2020 or the final months of 2024.

The average age of clients with STAR clinical encounters during this period was 44. Slightly more than half of clinical encounter clients were white, and slightly more than half were men. Almost half of all clients have experienced homelessness, and most clients had just one STAR clinical encounter. Almost three-quarters of clinical encounter clients have had prior interactions with police, most commonly through police contacts, which include less serious interactions with law enforcement such as field interviews and general occurrences. More than a third of all individuals with STAR clinical encounters have had more serious law enforcement interactions including arrests and/or bookings (table 2).

A disproportionately larger share of STAR clinical encounter clients are Black or African American compared with the share of the population of Denver (21 percent of STAR clients vs. 9 percent of the city's population), while Hispanic and Latino/a/x clients make up a disproportionately smaller share (12 percent of STAR clients vs. 28 percent of the city's population). While we do not have the data to assess what causes these disparities—for example, socioeconomic, health, and cultural considerations are but a few of the potential explanatory factors—known barriers impacting Hispanic/Latino/a/x individuals' access to various types of programs and services include linguistic barriers and concerns around legal

documentation and associated repercussions. Understanding these challenges is critical to enhancing equitable access to STAR services.

TABLE 2

Characteristics of STAR Clinical Encounter Clients (N = 6,664)

Characteristic	*n*	Share
Demographics		
Age		
Under 18	524	8%
19-24	2,836	43%
25-44	1,978	30%
45-64	893	13%
65+	432	6%
Missing	1	
Gender		
Man	3,100	53%
Woman	2,592	45%
Prefer to self-describe	30	1%
Non-binary	84	1%
Missing	858	
Race/ethnicity		
White	3,026	52%
Black or African American	1,204	21%
Multiple races/other	678	12%
Hispanic, Latino/a/x	696	12%
AANHPI	104	2%
American Indian or Alaska Native	92	2%
Missing	864	
Housing status		
Ever experienced homelessness	2,932	44%
Clinical STAR van encounters		
1 encounter	5,068	76%
2–5 encounters	1,404	21%
6+ encounters	192	3%
Prior interactions with police		
Any police history	4,861	73%
Police contacts		
0	1,875	28%
1	767	12%

Characteristic	*n*	Share
2–5	1,644	25%
6+	2,387	36%
Arrests		
0	4,270	64%
1	737	11%
2–5	859	13%
6+	798	12%
Bookings		
0	4,557	68%
1	612	9%
2–5	879	13%
6+	616	9%

Source: Data on demographics, housing status, and clinical STAR van encounters are from WellPower. Data on prior interactions with police from Denver Department of Public Safety.
Notes: Data include STAR services from June 2020 (beginning of STAR program implementation) to September 2024. "Police contacts" include field interviews and general occurrences. AANHPI = Asian American, Native Hawaiian, or Pacific Islander. "Missing" categories are not included in calculations for total shares.

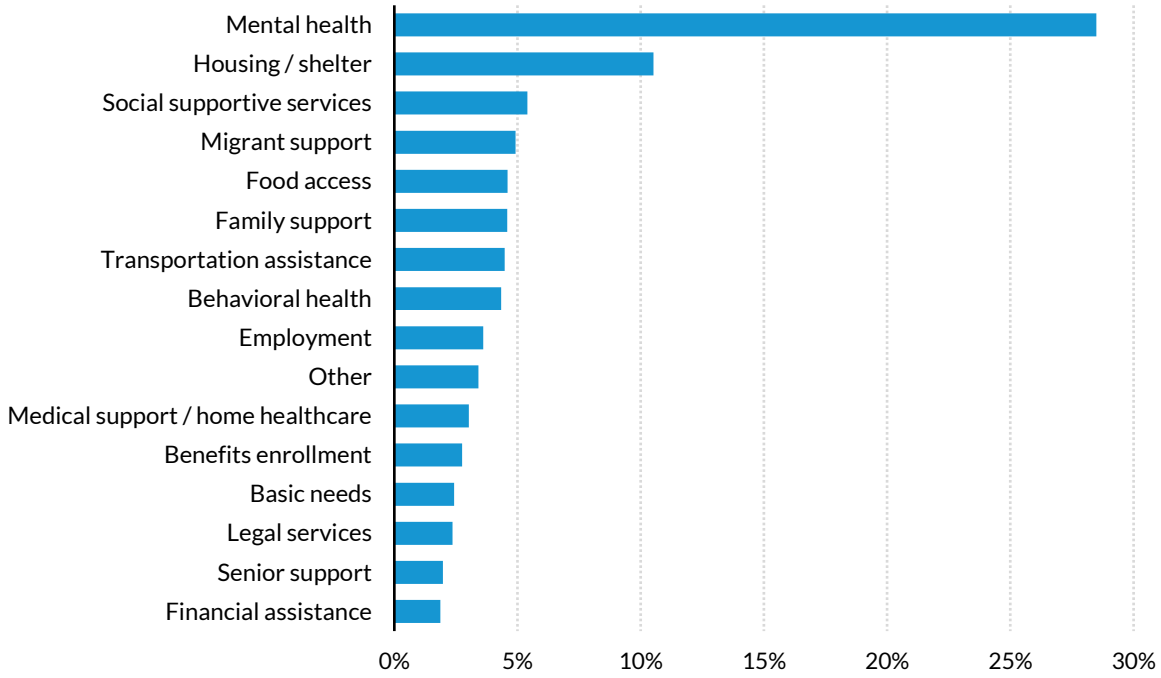
According to clinical encounter data, more than three-quarters of encounters reported mental health as a priority issue, and most encounters resulted in interventions such as general support/rapport building (91 percent of encounters), clinical assessment (50 percent of encounters), and/or community referral (38 percent of encounters). These interventions are not mutually exclusive because clients could receive multiple types of intervention during a clinical encounter. About 47 percent of clinical encounters were referred to a service provider and clients could be referred to multiple providers.

The STAR Community Partner Network began accepting referrals in April 2023. Over the first 1.5 years (April 2023 through September 2024), the STAR Community Partner Network received 1,357 referrals. Most referrals (98 percent) came through the STAR van teams (along with a few direct referrals from other community partners), which represents about 21 percent of all STAR van clinical encounters during the same period. About 29 percent of these referrals were for Hispanic or Latino clients, a notable difference compared with STAR van clinical encounters, where Hispanic/Latino clients account for only 12 percent of all individuals. These referrals represented 1,221 unique clients who engaged in almost 12,000 case management encounters with the STAR Community Partner Network, an average of 10 case management encounters per client. Most commonly, case management encounters focused on mental health, with housing/shelter related resources the second most common focus (figure 4).

FIGURE 4

Nearly One in Three Case Management Encounters Addresses Mental Health

STAR Community Partner Network case management encounters by focus area (N = 11,734)



URBAN INSTITUTE

Source: Servicios de La Raza (STAR Community Partner Network lead agency).
Notes: Data are from the period April 1, 2023–September 26, 2024. Case management service categories provided during less than three percent of encounters are omitted from this figure for brevity.

Outcomes

In table 3, we present results from a series of regression models that compare our propensity score-matched treatment (STAR clinical encounter) and comparison (police field interview) groups. We look at the effects of receiving a STAR clinical encounter in lieu of a police field interview on individuals' subsequent events—police contacts, arrests, bookings, and other STAR encounters—in the year following their first STAR clinical encounter. We explore the effects on continuous (count) outcomes, such as the number of arrests in the following year, as well as binary outcomes, such as whether an individual had *any* arrest in the following year. These approaches provide complementary pictures of the effects of STAR, with binary outcomes describing the share of people who, for example, avoided arrest in the following year, and with continuous outcomes describing the average number of arrests avoided.

TABLE 3

STAR Clinical Encounters Reduce Likelihood of Subsequent Police Contacts and Arrests

Changes in police contacts, bookings, and STAR encounters one year after a STAR clinical encounter

	Regression-Adjusted Means		
Event type in the year following an encounter	**Comparison group (police)**	**Treatment group (STAR)**	**Difference in means**
Binary outcomes			
Any police contacts	0.50	0.42	-0.08 *** ^^^
Any arrests	0.18	0.15	-0.02 * ^^
Any bookings	0.15	0.15	0
Any STAR encounters	0.03	0.15	0.11 *** ^^^
Continuous outcomes			
Number of police contacts	1.62	1.46	-0.16 *
Number of arrests	0.42	0.39	-0.03
Number of bookings	0.29	0.30	0.02
Number of STAR encounters	0.05	0.36	0.31 *** ^

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on STAR clinical encounters come from WellPower.

Notes: The STAR encounter sample is defined as first STAR clinical encounters between June 1, 2020, and July 30, 2024. The comparison encounter sample is defined as any field interview with a nature code similar to that of a STAR-eligible call for service over the same time period who had never had a prior STAR clinical encounter. The sample size is 4,167 in the STAR encounter group and 87,826 in the police encounter comparison group. The comparison group was identified via propensity score matching. The results were estimated using ordinary least squares for continuous outcomes and logit for binary outcomes. These regression-adjusted models included the following control measures: age, race/ethnicity, encounter reason, and prior homelessness. In addition, each regression controlled for the pre-enrollment measure of the outcome. Outcomes included arrests, police contacts (defined as field interviews and general occurrences), bookings, and STAR clinical encounters.

* = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

We find that people who had a STAR clinical encounter had a statistically significant reduction in the likelihood of having any police contacts and any arrests in the year following that STAR encounter (table 3; refer to box 1 for guidance on interpreting statistical results). Half of all people who had a police encounter experienced at least one police contact during the subsequent year, compared with 42 percent of those who had a STAR clinical encounter. And 18 percent of all people who had a police encounter experienced at least one arrest during the subsequent year, compared with 15 percent of those who had a STAR clinical encounter. Both of these changes correspond to a 16 percent reduction in the probability of any police contact or any arrest in the year following a STAR clinical encounter. In other words, about one out of every six people who would have had a subsequent police contact or a subsequent arrest avoided that outcome. The associated effect size (annotated in table 3 with a "^" symbol) indicates that the reduction is nontrivial. We also found that people who had a STAR clinical

encounter were five times more likely to have a subsequent STAR encounter during the following year, and this increase was statistically significant. We did not find a significant change in subsequent bookings.

We also evaluated effects on the *number* of subsequent criminal-legal interactions. We found that STAR encounters were associated with statistically significant reductions in the number of police contacts during the subsequent year. STAR encounters were also associated with a statistically significant increase in the number of subsequent STAR encounters. Although these changes represent small reductions at the individual level, they can become meaningful over time and when aggregated at the program level. For example, STAR encounters reduced police contacts by one-sixth of a police contact per client per year. Given that the program averaged 1,265 first-time clinical encounters annually during the study period, that translates to an estimated 202 fewer police contacts per year at the program level.[5] In addition to improving outcomes and increasing access to appropriate services for clients, such reductions can generate substantial cost offsets by avoiding more expensive criminal justice interactions. Some binary outcomes were statistically significant and had moderate or large effect sizes, but the counterpart continuous outcomes were not statistically significant or did not have noted effect sizes. Take, for example, the binary outcome for arrests (-0.02 * ^^) compared with the continuous outcome (-0.03). These results suggest that while the likelihood of having any arrest in the following year is significantly lower—and that this effect is substantial in size—the impacts on the total number of arrests (the continuous outcome) in the following year are not significant, nor are they substantial in size. One possible explanation for this is that a limited number of individuals account for a large share of all arrests, so while the probability of any given person being arrested decreased meaningfully, the total number of arrests in the year following the focal encounter was only marginally smaller.

We also examined whether STAR clinical encounters had differential impacts across different types of clients. (See Appendix B for all regression models.) The results in table 4 show changes in one key outcome of interest—the number of subsequent police contacts—during the one year following a STAR clinical encounter across different client subpopulations. While clinical encounters reduce the number of subsequent police contacts for each subpopulation, some subpopulations experienced greater impacts from receiving a clinical encounter than others. Most notably, we found substantially larger impacts for clients who had ever experienced homelessness, as documented by police and reflected in the years of public safety data used for this study. The next largest difference was for clients with one or more prior arrests, followed by clients under age 40. These results suggest that STAR van encounters

may be working by interrupting cycles of criminal justice involvement or by meeting the particular needs of younger clients to help divert future criminal justice involvement.

TABLE 4

STAR Clinical Encounters Are Particularly Impactful for Individuals with Histories of Homelessness

Changes in police contacts one year after a STAR clinical encounter, by subpopulation

	Sample Sizes		Regression-Adjusted Means		
Subpopulation	Treatment	Comparison	Comparison group (police)	Treatment group (STAR)	Difference in means
Clients who identify as white	2,248	39,355	1.54	1.41	-0.13
Clients who identify as a person of color	1,919	48,471	1.70	1.51	-0.18
Clients who have ever experienced homelessness	2,192	13,660	3.28	2.84	-0.44 ***
Clients who have never experienced homelessness	1,975	74,166	1.27	1.30	0.03
Clients under 40	1,873	50,782	1.66	1.44	-0.22 *
Clients 40 and over	2,294	37,044	1.60	1.44	-0.16 *
Clients with one or more prior arrests	906	20,453	3.17	2.77	-0.41 *
Clients with no prior arrests	3,261	67,373	1.17	1.04	-0.12

Sources: Data on police contacts come from the Denver Department of Public Safety. Data on STAR client characteristics come from WellPower.

Notes: The STAR encounter sample is defined as first-time clinical encounters (only one per client) between June 1, 2020, and July 30, 2024, subset to the specified subpopulation. The comparison police encounter sample is defined as field interviews only (the lowest level of police encounters where no crime or violence has taken place) over the same time period. Each subpopulation comparison group was identified via propensity score matching. Results were estimated using ordinary least squares for continuous outcomes and logistic regression for binary outcomes. These regression-adjusted models included the following control measures: encounter year, encounter reason, client age, client race/ethnicity, and client history of homelessness. In addition, each regression controlled for pre-enrollment measures of criminal justice encounters. The outcome is subsequent police contacts (defined as field interviews and general occurrences).

* = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

Cost

STAR program costs represent the two main components of the program—the STAR van teams staffed by WellPower and Denver Health and Hospital Authority and the STAR Community Partner Network managed by Servicios de la Raza. Denver Department of Public Health and Environment and Denver 9-1-1 also provide in-kind resources through key personnel for program leadership and operations, and the STAR van teams leverage Medicaid reimbursement for covered services.

In 2023, when STAR was in full implementation and for which we have complete cost information, the total annual cost for the van teams was $2.49 million, and $542,000 of that was the one-time capital costs of new van purchases and upfitting. Without that, the cost of the STAR van teams totaled close to $2 million. In 2023, the cost for the Community Partner Network (which began accepting referrals in April 2023) was $1.92 million, bringing the total cost for these two program components to $4.4 million, or $3.86 million without one-time capital costs. It is important to note these cost estimates do not reflect Medicaid reimbursement for covered services, which represented 23 percent of the van teams' total documented costs in 2023. In addition, the Denver Department of Public Health and Environment provided two staff positions plus leadership time, and Denver 9-1-1 estimated almost $460,000 of in-kind support for leadership time, dispatch time, and training development and implementation; these in-kind costs are not included in the total program cost estimates.

In 2023, STAR van teams responded to 8,215 calls for service. A simple estimate of the cost per response calculated as STAR van team costs only divided by STAR van team responses is $237 per van team response. Using the same approach, a simple estimate of the cost per response for both the van team and average follow up services is $470 per response. These are imperfect estimates, particularly because data on Community Provider Network referrals and case management clients per month is limited during this time period. But it provides a back-of-the-envelope estimate that can be helpful for understanding the financial resources needed for the STAR program as it was implemented in 2023.

For comparison, our previous research in Denver estimated that an average police contact cost $76; an average arrest, including a 9-1-1 call and court appearance, cost $1,011; and an average jail day for someone with serious mental health challenges cost $239 (Gillespie et al. 2021).[6] STAR is a resource-intensive intervention and is designed to do much more than, for example, a police field interview, which is a low-level police encounter that would likely happen for many STAR-eligible calls for service in the absence of STAR. Unlike police contacts, STAR is designed to de-escalate and provide focused assessment and intervention on scene with STAR van teams, then connect clients to follow-up services in the community to increase stability. It makes sense that a clinical van encounter or van encounter plus follow-up services would cost more than one-time, brief police contacts.

At the same time, STAR responses avoid more serious and costly police interactions and criminal justice outcomes down the road. In addition to saving at least $76 for each STAR response that avoids a police contact, the outcomes discussed above (table 3) showed that one in six people who encounter STAR will then avoid an arrest in the following year, for a total direct cost avoidance of $1,087 per avoided arrest. However, this narrow example doesn't account for all the other potential outcomes associated with avoiding an arrest that we haven't measured in this study, such as the value of health

and well-being benefits. Thus, one important part of the cost-effectiveness argument for STAR is that it not only provides better outcomes for clients but also avoids the significant costs associated with some police responses.

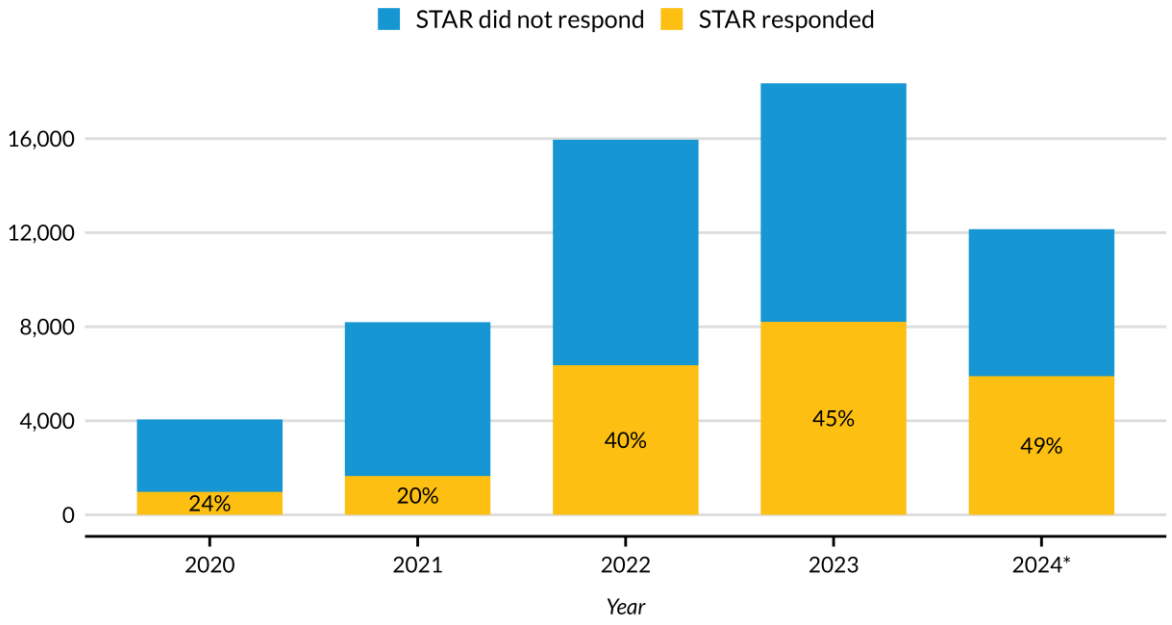
Scalability

Since STAR's inception, the number of eligible calls for STAR services has far outpaced the response capacity of STAR van teams. In the first year, the pilot program operated with one van unit—which constitutes one staffing team that can respond to calls for service—that responded to 28 percent of all calls for STAR services, including both STAR-eligible calls and calls not initially marked as STAR-eligible but to which STAR responded. By 2023, the program operated six van units (adding a seventh unit at the very end of the year) that responded to 45 percent of all calls for STAR services (figure 5).

FIGURE 5

There Is Demand for Thousands of Additional STAR Responses Each Year

Calls for service and STAR responses by STAR response status (N = 58,719)



URBAN INSTITUTE

Source: Authors' analyses of call-for-service data and STAR responses provided by the Denver Department of Public Safety.
Notes: Calls for service include STAR-eligible calls and calls that were not initially flagged as STAR-eligible that STAR responded to.
*Data include STAR services from June 2020 (beginning of STAR program implementation) to September 2024. As a result, data for 2024 does not cover the entire year.

Estimating Capacity Needed to Meet Demand

Using the reported demand for STAR in 2023 (the most recent full year of data), there were 15,125 STAR-eligible calls and 3,233 other STAR responses (typically requested by other agencies) for a total of 18,358 calls for STAR van services (table 5). In 2023, six STAR units responded to 8,215 calls for service (45 percent); however, not all units operated every day. Based on the response data, an average STAR unit operated about 71 percent of days in 2023, or an average of five days a week. The average STAR unit responded to approximately 5.3 calls per day during one 12-hour shift (down about 20 percent from the pilot year, when the average unit responded to roughly 6.6 calls per day). The average unit's first assigned call and last assigned call in 2023 spanned approximately 7.7 hours, an increase from 5.7 hours during the pilot year. We don't know from the data what might be driving the number of responses or length of time for assigned calls from year to year. Assuming each STAR van unit averaged about 5.3 calls per day and the program operated all six units concurrently each day for seven days a week, the program's maximum potential response volume would have been about 11,600 van responses in 2023. That level of operating units would have increased the share of all calls for STAR services that received a STAR response to 63 percent.

The key to meeting more demand for STAR is operating more van units for more days a week. Looking ahead, if the program could maintain sufficient staffing and infrastructure to operate 9 or 10 van units concurrently each day (one shift per unit), the program could respond to an additional 20 calls per day and theoretically meet its full demand. This assumes key factors remain constant with 2023 trends, including the average number of responses per unit per day and the total volume of calls for STAR services. Assuming a staffing team of two can work three shifts a week, about 22 staffing teams would be needed to operate 9 or 10 van units concurrently on each day of the week. If trends changed—for example, if other agencies directly requested STAR services more often (currently these requests represent 18 percent of all STAR responses)—or if the number of responses per unit per day changed, this would affect the number of units and staffing teams needed to meet the demand for services.

TABLE 5

Calls for STAR Van Services and Van Unit Capacity in 2023

STAR calls and responses	Statistic
Annual calls for STAR services	**18,358**
STAR-eligible calls	15,125
Other calls not initially flagged for STAR	3,233
Annual STAR responses	**8,215**
To STAR-eligible calls	4,982
To other calls not initially flagged for STAR	3,233
Annual share of calls for STAR services that received STAR responses	45%
Average responses per day	5.3
Units needed to respond to all calls for STAR services	9.5

Source: Authors' analyses of call-for-service data provided by the Denver Department of Public Safety.
Notes: *Average days per unit per week is calculated by taking the combined number of days each STAR unit responded to a call for service (1,551) and dividing by the number of units (6) to get the average days each unit responds per year (258.5). 258.5 days is 71 percent of 365 days in the year, which translates to 5 out of 7 days a week.

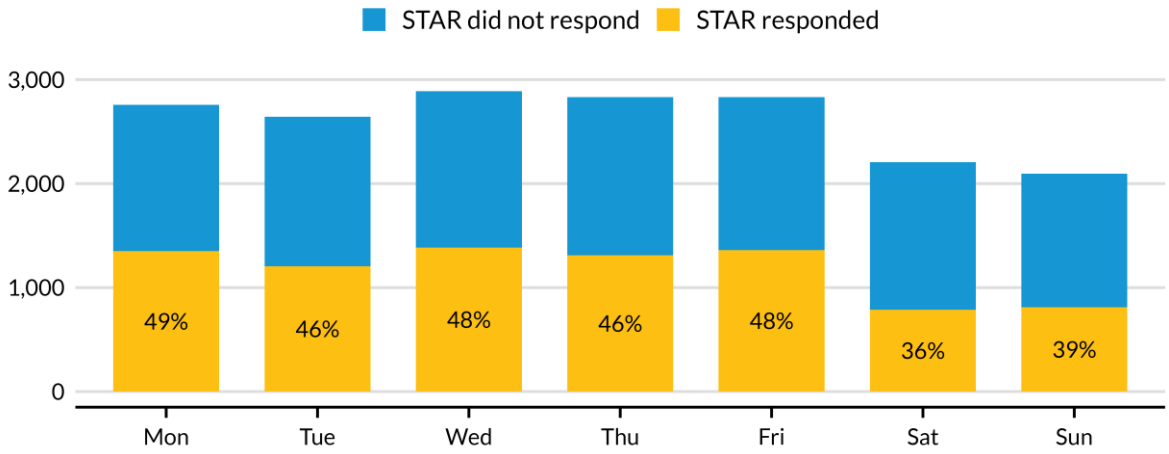
Considering Strategies to Meet Demand

Data highlight several ways more STAR units could respond to more of the calls for STAR services. In terms of days of the week, the share of calls for STAR services that get a STAR response is lowest on the weekend, at slightly under 40 percent of all calls compared with almost 50 percent of all calls during the week (figure 6).

FIGURE 6

STAR Response Rates Are Slightly Lower on Weekends

Calls for service and STAR responses by day of the week, 2023 (N = 18,253)



URBAN INSTITUTE

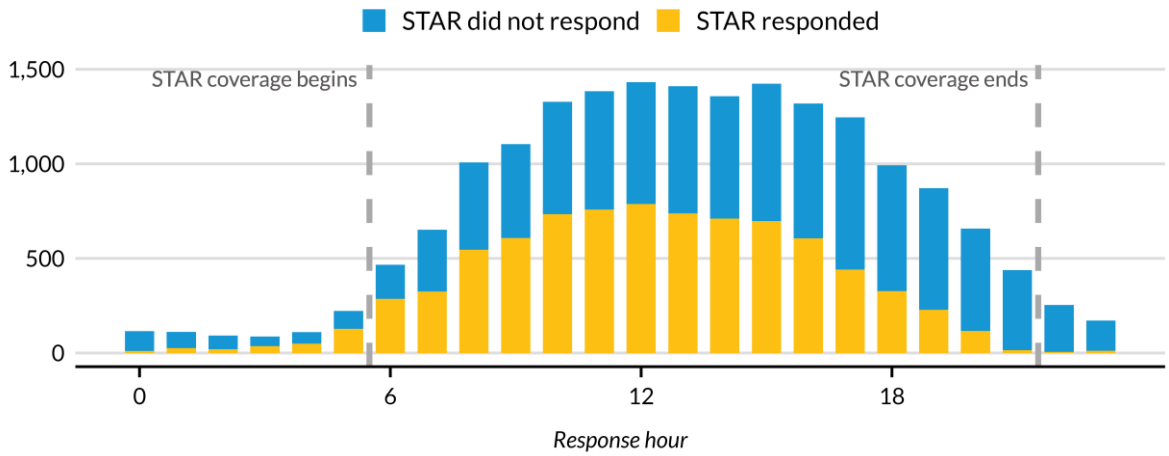
Source: Authors' analyses of call-for-service data provided by the Denver Department of Public Safety. Calls for service include STAR-eligible calls and calls that were not initially flagged as STAR-eligible that STAR responded to.
Notes: Call-for-service data are for calendar year 2023.

In terms of hours of the day, most calls for STAR services are received during current STAR operating hours. However, the share of calls for STAR services that get a STAR response is lowest during the later operating hours between 5:00 p.m. and 10:00 p.m. (figure 7).

FIGURE 7

Most Calls for Service Occur During STAR's Operating Hours with the Smallest Share of Responses in the Latest Hours

Calls for service and STAR responses by hour, 2023 (N = 18,253)



URBAN INSTITUTE

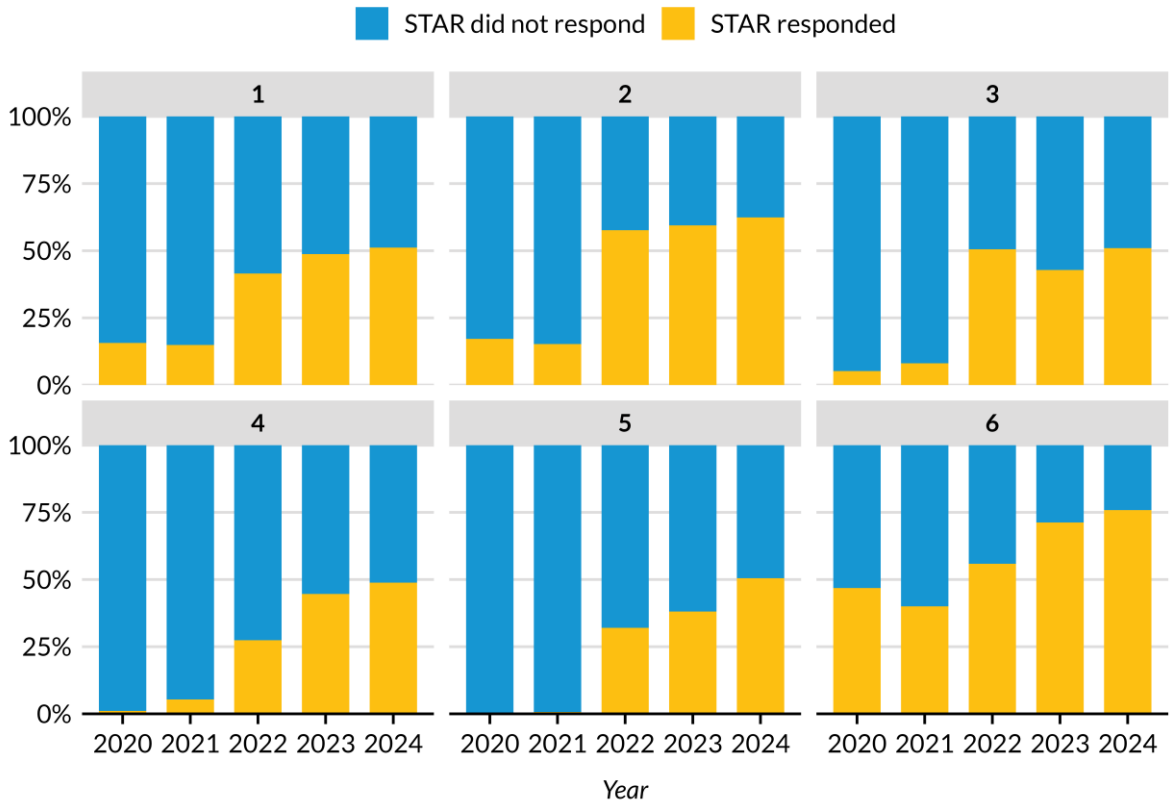
Source: Authors' analyses of call-for-service data provided by the Denver Department of Public Safety.
Notes: Call-for-service data are for calendar year 2023. Calls for service include STAR-eligible calls and calls that were not initially flagged as STAR-eligible that STAR responded to.

In terms of geographical location of STAR responses, although all police districts have more calls for STAR services than STAR responses (figure 8), the proportion of calls for STAR services that receive a STAR response varies. In 2024, STAR responded to about half of calls for STAR services in Police Districts 1, 3, 4, and 5. STAR vans responded to over 60 percent half of calls for STAR services in Police District 2 and over 75 percent of the demand in Police District 6.

FIGURE 8

STAR Responds to the Largest Share of Calls in Police District 6, which Encompasses Downtown Denver

STAR-eligible calls for service and STAR responses by police district (N = 58,719)



URBAN INSTITUTE

Source: Authors' analyses of call-for-service data provided by the Denver Department of Public Safety. Calls for service include STAR-eligible calls and calls that were not initially flagged as STAR-eligible that STAR responded to.
Notes: Call data represent the period between June 2020 and September 2024.

Another important consideration for scale is how much need for STAR services is not currently captured in call-for-service data, either because there are additional calls for service that may be appropriate for a STAR response but are not identified this way in the data or because there are additional people who need a STAR response but do not call 9-1-1 at all. As discussed in the implementation section above, individuals who identify as Hispanic or Latino/a/x are underrepresented in STAR clinical encounters, suggesting that unmet need for STAR services within this community may be substantially higher. In addition, as noted in the outcomes discussion above, individuals who have ever experienced homelessness appear to derive the greatest benefit from STAR responses in terms of

reductions in subsequent law enforcement encounters. Specific program recommendations are offered in the conclusion of this report.

Projecting Cost at Scale

The data described in the cost discussion above can also be used to project program costs and potential cost offsets if STAR were scaled to meet additional need. At a cost of $237 per van response, responding to all 18,358 calls for STAR services in 2023 could cost around $4.35 million for just the van team services (table 6). This same approach, when used to project costs for the full program at this scale, totals about $8.63 million including follow up services, although such estimates are limited in the absence of additional information about the cost per referral and case management services provided by the STAR Community Partner Network. At greater scale, the program would also be expected to realize larger cost offsets driven by the greater avoidance of more expensive criminal-legal system outcomes, as discussed above. Beyond those measurable savings, expanding STAR's reach would also increase the program's broader community impact by connecting more people to appropriate services for longer-term stability and well-being.

TABLE 6

STAR Program Costs at Scale, 2023

	Cost per response	Responses at scale	Cost at scale
Van teams only	$237	18,358	$4,350,846
Full program costs	$470	18,358	$8,628,260

Source: Authors' analyses of 2023 program cost and call-for-service data provided by the Denver Department of Public Safety.
Notes: Cost per response is calculated by dividing 2023 program costs (detailed above) by total STAR responses in 2023 including both STAR-eligible calls for service and STAR responses requested by other agencies. Responses at scale is calculated as all STAR-eligible calls for service and STAR responses requested by other agencies in 2023. Cost at scale is the cost per response multiplied by the number of responses at scale.

Co-Responder Findings

Co-responder programs—where police respond alongside mental health clinicians or crisis specialists—are among the most common models for crisis response, and these models have been implemented for much longer than civilian-led alternative response models. The research shows that co-responder programs are associated with improved immediate outcomes during crisis calls, particularly for mental health–related incidents (Dee and Pyne 2025). At the same time, research highlights clear limitations and gaps. Similar to alternative response program evaluations to date, most co-responder evaluations

have also focused on on-scene outcomes, such as same-day arrests, referrals, or call resolution, rather than post-response service engagement or long-term outcomes.

As alternative response programs grow, many communities are seeking to understand how alternative response and co-response work together to get the right response to the right call. Most traditional police–clinician co-responder models dispatch officers and clinicians together as routine practice, sharing on-scene roles in assessment, de-escalation, and disposition. While most civilian-led crisis response models, such as those in Denver, San Francisco, Albuquerque, Portland, and St. Petersburg, are designed to respond *instead* of police for eligible calls, law enforcement may request alternative response when those specific services are needed, and law enforcement may also be called to alternative response calls for safety backup in rare situations. Some programs are becoming more intentionally hybrid models—most notably CAHOOTS, which is clinician staffed but operates in close coordination with police, and Durham HEART.

Although CAHOOTS can respond independently, strong causal evidence recently showed that its largest reductions in arrests occurred when crisis teams and police co-responded, functioning as complements rather than substitutes. Calls in which CAHOOTS and police co-responded experienced substantially larger reductions in arrest risk than calls where CAHOOTS replaced police outright. This suggests that clinicians' presence helped de-escalate high-risk situations and provided alternatives to mental health or detoxification holds, while police retained authority for safety and enforcement when necessary. Co-response between police and CAHOOTS also increased access to medical services, including EMS and emergency room transport, indicating this model may help ensure crises are managed through medical pathways rather than coercive or custodial ones, supporting alignment of client needs to community services (Davis et al. 2025).

We began studying Denver's Co-Responder Program in 2022 alongside our evaluation of the STAR program. Our published work so far includes an implementation study focused on describing program successes and challenges as well as program recommendations based on staff interviews (Gillespie et al. 2023a). In this paper, we use program data and linked administrative data to characterize the scope and scale of the Co-Responder Program and to estimate its impacts on criminal justice outcomes for individuals in the year following a co-responder encounter.

Implementation

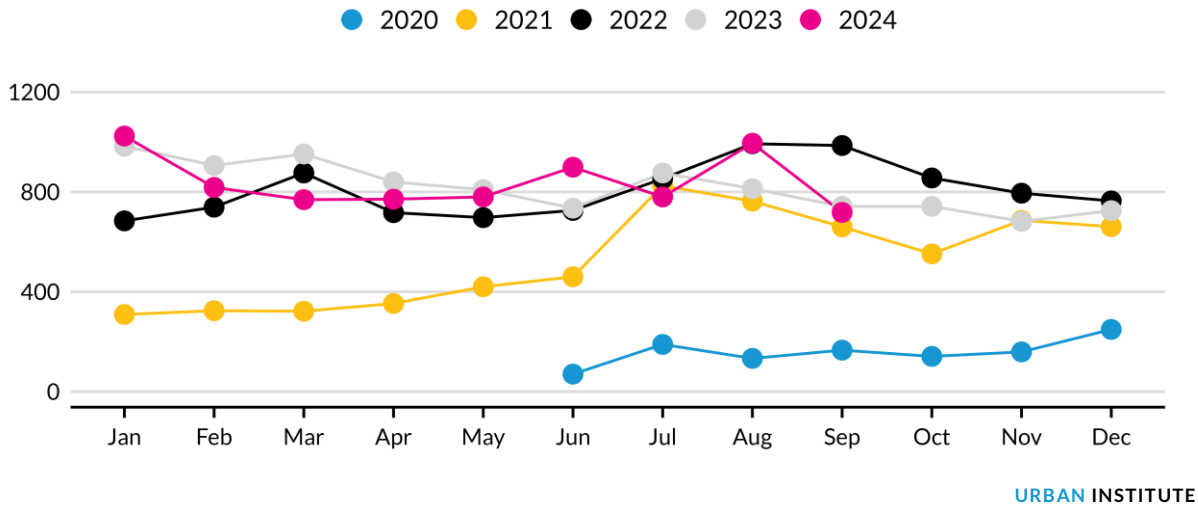
Denver's Co-Responder Program began in 2016 as a partnership between Denver Police Department (DPD) police officers and mental health clinicians. Co-responder clinicians have also been assigned to work with the Denver Fire Department, and over time the program has grown to include partnerships between clinicians and a diverse range of city agencies. By sending mental health experts to accompany police officers and others in first responder positions, the Co-Responder Program aims to divert individuals from excessive engagement with the criminal-legal and crisis-response systems by connecting them to culturally and geographically appropriate community support and services. Because co-responders accompany police officers, there are not equivalent "Co-Responder-eligible" calls-for service to which co-responders can respond. Rather, co-responders can theoretically respond to any call for service alongside a police officer, and in practice, co-responders often respond to calls based on whether their skills could be a good match for a given call for service. On the scene, the officer and clinician decide together how to respond with the goal of de-escalating the crisis and making a connection to follow-up services. After the initial call, co-responder clinicians may follow up themselves to provide referrals or may direct clients to a case manager, outreach case coordinator, or other appropriate services.

By spring 2018, the Co-Responder Program had expanded to all six police districts within Denver, with 24-hour coverage throughout the week comprising day shifts, swing shifts, and night shifts. By 2024, co-responders were responding to 800 to 1,000 calls for service each month (figure 9). For comparison, Denver 9-1-1 generates about 45,000 calls each month for police, fire, and medical services.[7] Among all co-responder calls, the most common nature codes were recorded as disturbance, welfare check, and suicide or self-harm.

FIGURE 9

Monthly Co-Responder Responses Have Held Fairly Steady since Summer 2021

Co-responder responses by month and year (N = 34,487)



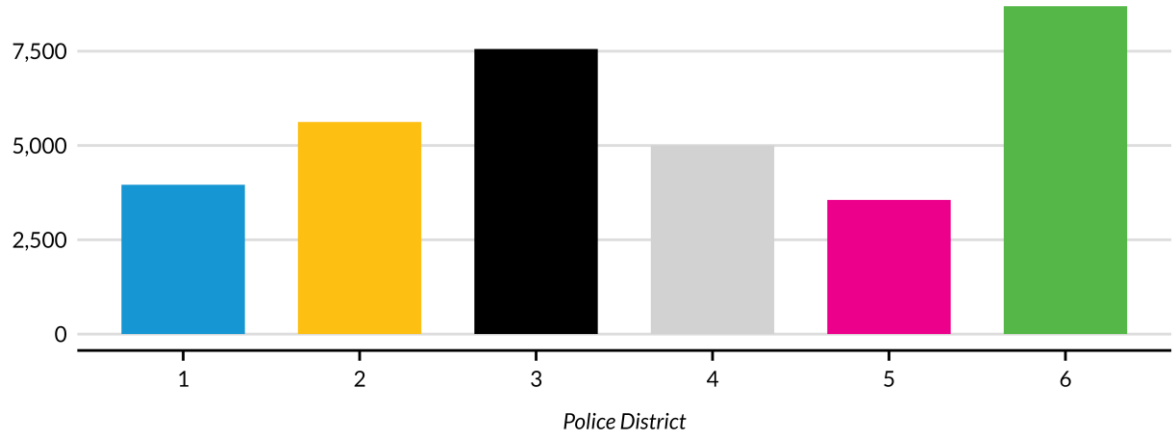
Source: Denver Department of Public Safety.
Note: Data for this figure include co-responder services from June 2020 to September 2024.

Altogether over the four and a half years—from June 2020 through September 2024—captured by our study, co-responders responded to 34,487 calls for service. Only about 1 percent (n = 359 calls) of those responses were for calls-for-service to which STAR also responded. Co-responders responded across all six police districts with the highest frequency of responses in Police District 6 (downtown) and Police District 3 (north-central and northeast, figure 10).

FIGURE 10

Police District 6, which Encompasses Downtown Denver, Has Received the Most Co-Responder Responses

Co-responder responses by police district (N = 34, 391)



URBAN INSTITUTE

Source: Denver Department of Public Safety.
Notes: Data for this figure include co-responder responses from June 2020 to September 2024, excluding those missing geographic information.

About one-third of all co-responder responses (10,183 responses) were documented by the clinician as a clinical encounter, intended to capture calls that result in a substantive engagement with the co-responder clinician. Unlike 9-1-1 calls for service and co-responder responses that do not rise to the level of a "clinical" response, clinical encounters include identifying information about individuals and their characteristics. There were 7,297 unique clients who had at least one clinical encounter with a co-responder team. Approximately one in five people with a clinical co-responder encounter had multiple clinical co-responder encounters, and a small subset of high-frequency clients (28 clients) had 10 or more clinical co-responder encounters. At the highest end, one client had 24 clinical co-responder encounters during this period, an average of almost one encounter every two months.

The average age of clients with co-responder clinical encounters during this period was 38, just over half were men and just over half were white (table 7). About 19 percent of clinical encounter clients were Black or African American, and another 13 percent were Hispanic or Latino/a/x. Black or African American clients are overrepresented in this client population compared with the 9 percent of Denver's full population that identifies as Black. Hispanic and Latino/a/x clients in this client population are underrepresented in comparison with the 28 percent of Denver's full population that identifies as Hispanic. About one quarter of all clients have experienced homelessness, and most clients had just one co-responder clinical encounter. Eighty-five percent of clinical encounter clients have had prior

interactions with police, most commonly through police contacts, which include less serious interactions with law enforcement. Forty-two percent of clinical encounter clients have had more serious law enforcement interactions including arrests and/or bookings.

TABLE 7

Characteristics of Co-Responder Clinical Encounter Clients (*N* = 7,297)

Characteristic	*n*	Share
Demographics		
Age		Mean age = 38
Under 18	911	12%
19-24	718	10%
25-44	3,344	46%
45-64	1,718	24%
65+	605	8%
Missing	1	
Gender		
Man	3,744	54%
Woman	3,143	45%
Prefer to self-describe	40	1%
Non-binary	69	1%
Missing	301	
Race/ethnicity		
White	3,221	51%
Black or African American	1,227	19%
Multiple races/other	851	13%
Hispanic, Latino/a/x	805	13%
AANHPI	127	2%
American Indian or Alaska Native	80	1%
Missing	986	
Housing status		
Ever experienced homelessness	2,017	28%
Clinical co-responder encounters		
1 encounter	5,866	80%
2–5 encounters	1,325	18%
6+ encounters	106	2%
Prior interactions with police		
Any police history	6,178	85%
Police contacts		
0	1,168	16%
1	879	12%
2–5	2,086	29%
6+	3,164	43%
Arrests		
0	4,372	60%
1	885	12%
2–5	1,186	16%
6+	872	12%

Characteristic	*n*	Share
Bookings		
0	4,774	65%
1	782	11%
2–5	1,030	14%
6+	711	10%

Sources: Data on demographics, housing status, and clinical encounters are from WellPower. Data on prior interactions with police are from Denver Department of Public Safety.
Notes: Data include services from June 2020 (beginning of STAR program implementation) to September 2024. "Police contacts" include field interviews and general occurrences. AANHPI = Asian American, Native Hawaiian, or Pacific Islander. "Missing" categories are not included in calculations for total share.

More than 85 percent of clinical encounters reported mental health as a priority issue, and more than 25 percent reported substance use as a priority issue. Only a small share of encounters reported aging, physical health, or intellectual disability as priority issues. Most encounters resulted in interventions such as general support/rapport building (65 percent of encounters) and clinical assessment (47 percent of encounters). Priority issues and interventions are not mutually exclusive—clients could receive multiple types of intervention during a clinical encounter. About 40 percent of clinical encounters documented a referral to a service provider, and most referrals were to WellPower, a community mental health center that both serves as the co-responder partner with police and provides other services in the community.

Outcomes

For people who had a clinical co-responder encounter, we found a statistically significant reduction in the likelihood of any arrests in the year following that clinical encounter (table 8; refer to Box 1 for guidance on interpreting statistical significance and effect sizes). Thirty-three percent of people who had a police-only encounter experienced at least one arrest during the subsequent year, compared with 28 percent of those who had a co-responder clinical encounter. That corresponds to a 15 percent reduction in the probability of any arrest following a co-responder encounter. The associated effect size (denoted as "^^") indicates that this reduction is substantial in magnitude. We also found a reduction in the likelihood of other types of criminal justice encounters in the year following a co-responder encounter, including any police contacts and any bookings, though these reductions were not statistically significant. Finally, we found that people who had a co-responder encounter were five times more likely to have a subsequent clinical co-responder encounter during the following year, and this increase was statistically significant.

Another way to estimate the impact of the Co-Responder Program is to measure changes in the number of criminal justice interactions experienced by people who had a co-responder encounter or a police-only encounter. Among people who had a co-responder clinical encounter, we found statistically significant reductions in the number of police contacts, arrests, and bookings during the subsequent year, as well as a statistically significant increase in the number of co-responder encounters, compared with people who had a police-only encounter. Although these changes represent fractional reductions at the individual level, they can become meaningful over time and when aggregated at the program level. For example, the Co-Responder Program reduces police contacts by approximately one-sixth of a police contact per client per year. Given that the program averaged 2,263 clinical encounters annually during the study period, this translates to an estimated 362 fewer police contacts per year at the program level.

TABLE 8

Co-Responder Clinical Encounters Reduce Likelihood of Subsequent Arrests

Changes in Police Contacts, Bookings, and Co-Responder Encounters One Year After a Co-Responder Clinical Encounter

	Regression-Adjusted Means		
Event type in the year following an encounter	**Comparison group (police)**	**Treatment group (co-responder)**	**Difference in means**
Binary outcomes			
Any police contacts	0.67	0.65	-0.02 ^
Any arrests	0.33	0.28	-0.04 *** ^^
Any bookings	0.35	0.34	-0.01
Any co-responder encounters	0.03	0.15	0.11 *** ^^^
Continuous outcomes			
Number of police contacts	3.32	3.15	-0.16 *
Number of arrests	1.06	0.96	-0.1 ***
Number of bookings	0.82	0.75	-0.06 *
Number of co-responder encounters	0.07	0.22	0.15 ***

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on co-responder encounters come from WellPower.

Notes: The co-responder encounter sample is defined as co-responder clinical encounters between June 1, 2020, and July 30, 2024. The comparison police encounter sample is defined as any type of police response (field interview, general occurrence, or arrest) over the same time period. The sample size is 2,253 in the co-responder encounter group and 268,832 in the police encounter comparison group. The comparison group was identified via propensity score matching. Results were estimated using ordinary least squares for continuous outcomes and logistic regression for binary outcomes. These regression-adjusted models included the following control measures: encounter year, encounter reason, client age, client race/ethnicity, and client history of homelessness. In addition, each regression controlled for the pre-enrollment measure of the outcome. Outcomes included subsequent arrests, police contacts (defined as field interviews and general occurrences), bookings, and clinical co-responder encounters.

* = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .>.8.

We also examined whether the Co-Responder Program had differential impacts across different types of clients. (See appendix B for all regression models.) Most notably, we found substantially larger differences for clients who had previously experienced homelessness. Across all outcomes, the magnitude of change for this subgroup was two to three times larger than the changes observed for the full client population discussed above (table 9). Because arrest rates among clients experiencing homelessness were two to three times higher than those of the full client population, the impact of the program translates into larger absolute reductions in arrests for this subgroup.

Program data indicate the Co-Responder Program had clinical encounters with an average of 362 clients each year who had experienced homelessness. Based on our estimates, the program's impact among this subgroup resulted in approximately 206 fewer police contacts (-0.57 per person), 134 fewer arrests (-0.37 per person), and 80 fewer bookings (-0.22 per person) each year. In addition to improving outcomes and increasing access to appropriate services for clients, these reductions can generate substantial cost offsets by avoiding more expensive criminal justice interactions.

TABLE 9

Co-Responder Clinical Encounters Are Particularly Impactful for Individuals with Histories of Homelessness

Changes in police contacts, arrests, bookings, and Co-Responder encounters one year after a Co-Responder clinical encounter for people with histories of homelessness

	Regression-Adjusted Means		
Event type in the year following an encounter	**Comparison group (police)**	**Treatment group (co-responder)**	**Difference in means**
Binary outcomes			
Any police contacts	0.78	0.73	-0.05 * ^^
Any arrests	0.60	0.50	-0.09 *** ^^^
Any bookings	0.60	0.58	-0.03 ^
Any criminal justice encounters	0.87	0.81	-0.05 *** ^^^
Any co-responder encounters	0.06	0.18	0.13 *** ^^^
Continuous outcomes			
Number of police contacts	4.86	4.30	-0.57 ***
Number of arrests	2.66	2.28	-0.37 ***
Number of bookings	1.78	1.56	-0.22 ***
Number of criminal justice encounters	53.57	57.21	3.64
Number of co-responder encounters	0.12	0.27	0.14 ***

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on co-responder encounters come from WellPower.

Notes: This analysis for clients who have experienced homelessness includes encounters with people who have ever been flagged by police or WellPower clinicians as experiencing homelessness prior to the focal encounter, as recorded in the years of Denver Department of Public Safety data we used for the study. The co-responder encounter sample is defined as co-responder clinical encounters between June 1, 2020, and July 30, 2024. The comparison police encounter sample is defined as any type of police response (field interview, general occurrence, or arrest) over the same time period. For this sub-analysis, the sample size is 750 in the co-responder encounter group and 85,725 in the police encounter comparison group. The comparison group was identified

via propensity score matching. Results were estimated using ordinary least squares for continuous outcomes and logistic regression for binary outcomes. These regression-adjusted models included the following control measures: encounter year, encounter reason, client age, client race, and client history of homelessness. In addition, each regression controlled for the pre-enrollment measures of criminal justice encounters. Outcomes included subsequent arrests, police contacts (defined as field interviews and general occurrences), bookings, and clinical co-responder encounters.
* = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .>.8.

Conclusion

This evaluation of Denver's STAR and Co-Responder programs provides compelling evidence for the broader alternative crisis response and community responder fields. The findings reinforce prior research, including evidence of amplified benefits for clients with high baseline system involvement and the distinct yet complementary roles of civilian-led and police–clinician models. To our knowledge, the paper also offers the first analysis of program impacts on individual outcomes during the year following alternative response or co-response. Finally, the study identifies important directions for future research to address analytical limitations and offers actionable recommendations for program leaders to strengthen implementation and maximize program impact relative to cost.

Limitations and Future Research

In earlier sections, we provided a detailed discussion of the assumptions and limitations inherent in our analytic approach, particularly with respect to estimating the impacts of STAR van clinical encounters. Limitations in computer-aided dispatch (CAD) data—including evolving protocols over the study period and limited identifiers for linking unique people across data systems—constrain the precision with which calls, encounters, and outcomes can be classified, compared, and tracked longitudinally. Although we employ propensity score matching and regression adjustment to strengthen causal inference, our estimates may still be influenced by unobserved differences between people who receive a STAR or co-responder clinical encounter and those who receive a police-only response. These data limitations also mean this study measured only the impact of clinical encounters recorded by program staff and impeded our ability to capture unmet demand for services, including people who could have benefited from STAR or co-responder responses but were either not identified as eligible in the data or did not contact 9-1-1 at all. We were unable to examine follow-up referrals or service engagement, even though these were core components of the STAR and Co-Responder models and critical to understanding these programs and their impact. We were not able to conduct cost or scalability analyses for the Co-Responder Program; however, given the strong outcomes observed, these represent important areas for future research to inform ongoing resource allocation and policy decisions.

Finally, the study focused on criminal-legal outcomes that could be measured using existing administrative data provided by program partners and the City of Denver. We could not measure outcomes related to health and well-being or housing stability, for example. These challenges underscore the need for continued data integration, improved outcome measurement, and complementary qualitative and primary data collection to more fully assess the impacts of community responder models. Addressing such gaps is essential, as facilitating service connections is a primary objective of both the STAR program and the Co-Responder Program and a central mechanism through which community responder models are intended to improve outcomes for people and communities.

Recommendations

Although the STAR and Co-Responder programs differ in structure, scale, and operational roles, both are designed to shift responses to behavioral health–related 9-1-1 calls away from traditional policing and toward clinical, service-oriented interventions. Each program serves people experiencing mental health crises, substance use challenges, or related needs and seeks to reduce reliance on arrest and detention while improving outcomes for clients and communities. In both programs, clinical encounters are associated with reductions in subsequent criminal justice involvement, including fewer police contacts and arrests, particularly among individuals with high levels of interaction with law enforcement, such as those who have experienced homelessness. Even modest per person reductions in criminal justice outcomes translate into meaningful impacts when aggregated at the program level. These results should be interpreted within the context of the research limitations discussed above.

The study identifies several opportunities to further broaden the reach and impact of both programs. First, both programs would benefit from deeper examination of service access within the Hispanic and Latino/a/x community, which is significantly underrepresented among program clients. Targeted analyses and community engagement to better understand the drivers of this underrepresentation could inform adjustments to outreach strategies, response protocols, and program design to ensure equitable access and impact across communities.

Second, the programs demonstrate substantially greater impacts among certain populations, particularly individuals who have experienced homelessness. Ensuring responses to eligible calls for people experiencing homelessness when possible could amplify program outcomes and help diminish associated costs.

Third, both programs have the potential to increase both the total number of responses and the share of responses that are documented as clinical encounters. Increasing total responses connects more clients to services and also stands to increase the total benefits of each program, including avoiding subsequent criminal justice encounters, as we document here. The section on program scalability highlighted several possible strategies to meet more of the need for STAR services, including by increasing staffing levels and the number of concurrent van units operating each day, which are the primary constraints on responding to additional calls. The program may also consider strategies to increase the average number of responses per unit per day, if possible and appropriate. Meanwhile, increasing documentation of encounters will better position both programs to describe the full scope of the services they provide and their associated outcomes, and more robust data can also better support ongoing program improvement efforts. It remains unclear why many encounters are not treated as clinical encounters and whether that reflects limitations in data collection, staffing capacity, or other operational constraints.

An additional finding common to both the STAR and the Co-Responder programs is that experiencing a clinical encounter substantially increases the likelihood of having another clinical encounter in the subsequent year. This pattern may be client driven, as individuals who have previously engaged with STAR may better understand how to access the program's services through 9-1-1 or other pathways. It may also be program driven, if call takers are more likely to assign STAR or a co-responder when prior responses to the same individual or location are visible in dispatch systems. Given the strong outcomes associated with clinical encounters in both programs, further examination of what predicts repeat clinical encounters could enhance program implementation and impact. Additionally, data showed that more than 1,700 individuals had clinical encounters with *both* STAR and the Co-Responder programs during the study period. Though we do not look at the combined impacts of these programs here, this could also be a valuable area of future inquiry and potential program coordination.

Finally, both programs would benefit from investments in data integration and service-tracking infrastructure to better document follow-up referrals, service engagement, and longer-term outcomes. Doing so would enable a more complete understanding of each program's impact across the full service continuum—from the initial call for service, to the clinical encounter, to referral and follow-up services. Service connection is central to both programs' theories of change, yet current data collection and data sharing practices constrain evaluation primarily to criminal justice outcomes. Improved data systems would strengthen program management and continuous quality improvement efforts, clarify true program impacts, and support future cost-effectiveness analyses.

Taken together, these findings suggest that the STAR and Co-Responder programs function as complementary components of Denver's community responder system, serving overlapping populations through distinct pathways. STAR expands the city's capacity to respond to some types of crises without police involvement, while the Co-Responder Program improves outcomes in situations where police response remains necessary. Their combined impact appears greatest for individuals with the highest levels of system involvement, underscoring the value of a diversified system that aligns the best response to each crisis to deliver the best outcomes for both clients and the community.

Appendix A. Methods

Implementation, Cost, and Scalability Findings

For these components of the study, our goal was to describe the ways in which the STAR and Co-Responder programs were implemented by examining when and where those programs were responding and what types of calls they were responding to. We also examined how much the STAR program cost to implement and what may be necessary for STAR to scale to meet its full demand.

Data

There were four primary data sources for these components of the study: 9-1-1 call data from the Denver Department of Public Safety (DOS), STAR and co-responder clinical encounter data from WellPower, data on police encounters from DOS, and data on program costs from Denver Department of Public Health and Environment.

Denver 9-1-1 Call Data

These data represent 9-1-1 calls for service that were either STAR-eligible or had a recorded STAR or co-responder response. Each record represents a single incident. There are no identifiers present in these data to identify unique individuals across calls. The data contain information such as the date and location of each call, the call reason, whether and what type of vehicle was dispatched, time of arrival and clearance of any vehicles, and whether that call was deemed as STAR-eligible. The data span from the start of the STAR program on June 1, 2020, through September 26, 2024. We use the data to calculate where and when STAR vehicles responded, what types of calls they responded to, and the share of STAR-eligible calls with responses.

WellPower Data

These data represent clinical encounters with clients as documented by STAR van teams and co-responders. Each record represents a single encounter; individuals reflected in these data could have one or more encounter. The data span from the start of the STAR program on June 1, 2020, through to September 26, 2024. While these data capture many aspects of clinical encounters, we rely on a few key attributes for this analysis, including: the date of the encounter, an anonymous identifier for the client

(for linking to DOS data), the reason code associated with the 9-1-1 call for service, whether the client was experiencing homelessness, and the client's demographic characteristics.

DOS Data

These data describe various types and attributes of police encounters and the individuals involved, including field interviews, general occurrences, arrests, and bookings. The data span from January 1, 2016, through August 24, 2025. We aggregate these data to create an encounter-level dataset such that each record reflects a field interview, general occurrence, arrest, or booking. DOS data also provide attributes describing the individual's race and age, along with the reason for the encounter.

Denver Department of Public Health and Environment Data

These data present STAR program costs by year across program partners including Denver Department of Public Health and Environment, Denver Health, WellPower, Servicios de la Raza, and other costs such as in-kind resources and Medicaid reimbursement. We used 2023 data for the cost analysis.

Causal Assessment of STAR and Co-Responder Outcomes

Our goal was to describe the effects of receiving a STAR van encounter or co-responder encounter in lieu of a police response on recipients' criminal justice outcomes: number of arrests, number of police contacts (general occurrences, field interviews), number of days in jail, and number of bookings. We implemented a propensity score matching model to estimate the effects of each program, separately, on criminal justice outcomes.

Data

There were two primary data sources for the assessment of STAR outcomes: STAR and co-responder clinical encounter data from WellPower and police contact–related data from DOS.

WellPower Data

These data represent clinical encounters between STAR van teams or Co-Responder units and individual clients. A clinical encounter involves an interaction that leads to identifying and documenting a substantive engagement. These clinical encounters comprise only a portion of all STAR and co-responder responses, but because WellPower staff do not capture detailed information about other

responses not deemed clinical encounters, we only evaluate clinical encounters as part of our causal assessments. Each record in the data represents a single encounter; individuals reflected in these data could have one or more encounters. The data span from the start of the STAR program on June 1, 2020, through to September 26, 2024. While these data capture many aspects of clinical encounters, we rely on a few key attributes for this analysis, including the following: the date of the encounter, an anonymous identifier of the individual involved (for linking to DOS data), the reason (also referred to as a "nature" code) associated with the 9-1-1 call for service, whether the individual was experiencing homelessness, and the individual's race.

DOS Data

These data describe various types and attributes of police encounters and the individuals involved, including field interviews, general occurrences, arrests, and bookings. The data span from January 1, 2016, through August 24, 2025. We aggregate these data to create an encounter-level dataset such that each record reflects a field interview, general occurrence, arrest, or booking. DOS data also provide attributes describing the individual's race and age, along with the reason code for the encounter, which we use in identifying the comparison group (see below).

We combine the encounter-level data from DOS with the clinical encounter data from WellPower. From these data, we derive measures that quantify the numbers and types of police encounters before and after each clinical or police encounter. For example, in the WellPower data, we might have an observation documenting a STAR clinical encounter for an individual on June 1, 2022. Using our WellPower and DOS data, we then count the numbers of clinical encounters, field interviews, general occurrences, arrests, and bookings that same person experienced in the period preceding June 1, 2022, and separately, in the period following June 1, 2022. We use these pre-period attributes to identify the comparison group (see below), and we use the post-period attributes as our outcome measures.

Identification of the Comparison Group

While the approaches for our assessments of STAR and the Co-Responder Program outcomes were similar, there were differences in how we identified the comparison groups for each program.

STAR

When 9-1-1 receives a call for service, that call can be marked as "STAR-eligible" if it meets relevant criteria, such as entailing a very low risk of violence. For STAR-eligible calls, 9-1-1 dispatches either a STAR van or a police officer.[8] If a STAR van team is available, then a STAR van will be dispatched;

otherwise the police will be dispatched. In theory, all STAR-eligible calls that do not receive responses from STAR van teams should be similar to calls that receive a STAR van response.

The call-for-service data, which indicate STAR-eligible calls and responses, cannot be matched with the DOS data, which are critical data for the outcomes of interest. Instead, we use data from WellPower to identify the "treated" group—all individuals who have a first-time clinical encounter (the focal encounter for our analysis) with STAR. We also use data on police encounters from DOS to construct a comparison group comprising all field interviews with a reason code that best fits the types of events to which a STAR van team might respond for individuals who had not received a STAR van response before this focal police encounter for our analysis. We use field interviews because this is the most likely response when a STAR-eligible call for service does not receive a STAR van response. We used the following reason codes to subset field interviews down to those that were most likely to have been STAR-eligible: disturbance; acting suspicious; welfare check; around residence; and around business.

Due to these data limitations, our comparison group may include field interviews for calls for service that were not STAR-eligible or were not responses to 9-1-1 calls and may exclude other field interviews that were in fact STAR-eligible.

Co-Responder

We use data from WellPower to identify the "treated" group—all individuals who have a clinical encounter for the first time in the period included in our dataset (the focal encounter for our analysis) with a co-responder. Our approach to defining the comparison group for co-responder responses was simpler because co-responders can respond to any types of call for service (there is no "co-responder–eligible" designation process). Accordingly, we use DOS data to identify all field interviews, general occurrences, and arrests where the individual had no prior co-responder encounters and had at least one prior field interview that had a STAR-eligible–aligned nature code(disturbance; acting suspicious; welfare check (adult); around residence; or around business) before the focal police encounter. We take this last step because our full dataset only comprised individuals with at least one such field interview.

Propensity Score Estimation

To correct for potential bias in the comparison groups, we employed a propensity score matching (PSM) approach to match each "treated" focal encounter (a first-time STAR or co-responder clinical encounter) with a similar comparison focal encounter with the police (as defined above). Our approach to constructing our matched treated and comparison encounters is described below:

1. For each focal encounter—either a clinical encounter or comparison police encounter—we calculated attributes that we used to match on with the PSM model. These attributes include demographic characteristics (namely, race and age in the baseline year) as well as pretreatment measures that may relate to an individual's subsequent likelihood of experiencing one of our outcomes of interest. These other attributes included:
    a. The year of the encounter
    b. The reason code for the encounter (one of welfare, mental health, or other)
    c. Whether an individual had experienced homelessness prior to the encounter
    d. Whether the individual had experienced any arrests or jail days in the three years preceding the encounter, and
    e. Whether the individual had experienced 0, 1–2, or 3 or more field interviews (one measure) or general occurrences (a second measure).
2. Some individuals had multiple police encounters in a single day. Because some of those encounters may have been interrelated (e.g., a field interview that evolves into an arrest), we consolidated such observations by selecting the first police encounter that occurred on that day for that individual as the comparison focal encounter for purposes of matching.
3. We then implemented PSM models at the encounter level for encounters that met the following criteria:
    a. For the clinical encounter group:
        i. it was a clinical encounter;
        ii. it was an individual's first clinical encounter with either STAR or the Co-Responder Program (during the period included in our dataset); and
        iii. the encounter occurred after the start of the STAR program on June 1, 2020 (the start of the period included in our dataset).
    b. For the comparison group:
        i. the encounter was either a field interview for selected reasons defined above (for the evaluation of STAR) or a field interview, general occurrence, or arrest (for the evaluation of the Co-Responder Program);
        ii. the encounter was the first police encounter for an individual on the given day (for control cases);

iii. the encounter had a selected reason code as defined above (for the analysis of STAR only);

iv. the individual had no prior STAR or co-responder encounters; and

v. the encounter occurred after the start of the STAR program on June 1, 2020 (the start of the period included in our dataset).

4. We specified our primary PSM model using a generalized, or "quick," matching approach while measuring distance between observations using a general linear model to predict treatment as a function of the comparison attributes described under number 1 above. Quick matching pairs each treated case with one or more comparison cases, with comparison cases given varying weights reflecting their proximity or lack thereof to the treated case. We applied a caliper of 0.1 to ensure that the distance between treated cases and their comparisons was not overly large. We specified a series of alternate PSM models, including nearest neighbor models with and without exact matches and varying caliper distance, but the balance achieved with our quick matching approach (see tables A.1 and A.2) was by far the best.

For subgroup analyses only, small sample sizes and differences in classification across data sources required that we consolidate some variables into a reduced set of classes. These variable definitions included:

- **Race:** For subgroup analyses only, we retained two categories for race: (1) individuals who identify as white alone and (2) all other individuals, including those identifying as Black, Asian, Indigenous, Middle Eastern/North African, and multiple races or individuals for whom we had no race data. For individuals who reported their primary combined race-ethnicity as Hispanic, and for whom we had no other race data, we assigned them a race of white.

- **Age:** For subgroup analyses only, we defined two age categories: under 40 and 40 and older. This enabled us to retain large sample sizes in each category, which was important both for matching and for powering subsequent subgroup analyses.

TABLE A.1

Pre- and Post-Matching Bias of the Control and Treatment Groups: STAR

	Unmatched			Matched		
Variable	**Treated**	**Control**	**Std. diff.**	**Treated**	**Control**	**Std. diff.**
Year of encounter	2022.64	2021.89	0.72	2022.63	2022.65	-0.02
Ever unhoused	0.51	0.16	0.7	0.53	0.54	-0.03
Race — person of color	0.45	0.55	-0.2	0.46	0.46	0
Race — white	0.55	0.45	0.2	0.54	0.54	0
Age — <40	0.48	0.58	-0.19	0.45	0.45	0.01
Age — 40+	0.52	0.42	0.19	0.55	0.55	-0.01
Encounter reason — mental health	0.27	0	0.6	0	0	0
Encounter reason — welfare	0.11	0.86	-2.4	0.15	0.16	-0.03
Encounter reason — other	0.62	0.14	1	0.85	0.84	0.02
Prior three years						
Arrests, any	0.21	0.23	-0.05	0.22	0.22	-0.01
Jail day, any	0.18	0.2	-0.03	0.19	0.19	-0.01
Reports, 0	0.52	0.43	0.19	0.51	0.49	0.04
Reports, 1–2	0.27	0.31	-0.1	0.27	0.29	-0.03
Reports, 3 or more	0.21	0.26	-0.12	0.21	0.22	-0.02
Field interviews, 0	0.61	0.57	0.07	0.6	0.58	0.03
Field interviews, 1–2	0.24	0.26	-0.05	0.24	0.25	-0.02
Field interviews, 3 or more	0.16	0.17	-0.04	0.16	0.17	-0.02

Source: Authors' analyses of police contact and booking data provided by the City and County of Denver and STAR clinical encounter data provided by WellPower.

Notes: All columns refer to averages. For example, the first column represents the average of treated cases pre-matching. Std. diff. = standardized difference between the treatment and control groups.

TABLE A.2

Pre- and Post-Matching Bias of the Control and Treatment Groups: Co-Responder

	Unmatched			Matched		
Variable	**Treated**	**Control**	**Std. diff.**	**Treated**	**Control**	**Std. diff.**
Year of encounter	2021.72	2022	-0.22	2021.72	2021.75	-0.02
Ever unhoused	0.33	0.32	0.03	0.33	0.37	-0.07
Race — person of color	0.42	0.6	-0.35	0.42	0.43	-0.01
Race — white	0.58	0.4	0.35	0.58	0.57	0.01
Age — <40	0.59	0.58	0.02	0.59	0.6	0
Age — 40+	0.41	0.42	-0.02	0.41	0.4	0
Encounter reason — mental health	0.35	0.02	0.69	0.35	0.36	-0.02
Encounter reason — welfare	0.18	0.05	0.35	0.18	0.18	0.02
Encounter reason — other	0.47	0.93	-0.93	0.47	0.46	0.01
Prior three years						
Arrests, any	0.26	0.45	-0.44	0.26	0.31	-0.1
Jail day, any	0.23	0.4	-0.41	0.23	0.28	-0.11
Reports, 0	0.42	0.18	0.48	0.42	0.41	0.02
Reports, 1–2	0.33	0.27	0.13	0.33	0.32	0.02
Reports, 3 or more	0.26	0.55	-0.68	0.26	0.28	-0.04
Field interviews, 0	0.51	0.15	0.72	0.51	0.51	0.01
Field interviews, 1-–2	0.28	0.46	-0.38	0.28	0.27	0.03
Field interviews, 3 or more	0.2	0.39	-0.47	0.2	0.22	-0.05

Source: Authors' analyses of police contact and booking data provided by the City and County of Denver and STAR clinical encounter data provided by WellPower.

Notes: All columns refer to averages. For example, the first column represents the average of treated cases pre-matching. Std. diff. = standardized difference between the treatment and control groups.

TABLE A.3

Distribution of Propensity Scores: STAR

	Control cases	Treated cases
Sample	**87,826**	**4,167**
Propensity score		
0	4,249	4
0.1	75,827	823
0.2	4,106	827
0.3	1,954	578
0.4	855	390
0.5	285	192
0.6	223	250
0.7	170	348
0.8	105	415
0.9	52	340
0	4,249	4

Source: Authors' analyses of police contact and booking data provided by the City and County of Denver and STAR clinical encounter data provided by WellPower.

TABLE A.4

Distribution of Propensity Scores: Co-Responder

	Control cases	Treated cases
Sample	**268,832**	**2,253**
Propensity score		
0	37,138	61
0.1	229,192	1,456
0.2	1,741	215
0.3	446	136
0.4	181	111
0.5	102	139
0.6	27	78
0.7	3	41
>0.7	2	16

Source: Authors' analyses of police contact and booking data provided by the City and County of Denver and STAR clinical encounter data provided by WellPower.

Analysis

We run regressions to test the effects of clinical encounters in each program using the matched sample. Our primary outcomes of interest include:

- **Police contacts**, which include both field interviews and general occurrences, are lower-level police encounters in which no arrest is made;
- **Arrests,** include both custodial arrests—those where someone is booked into jail, also referred to as "bookings"—and non-custodial arrests, where an individual is cited and must appear in court but is not booked into jail;
- **Bookings** are custodial arrests, i.e., those arrests where the individual is booked into jail for the start of a jail stay;
- **Clinical encounters** for both STAR and the Co-Responder Program are the subset of encounters where the interaction between the responders and the individual rises to the level of a more substantive engagement and is documented by program staff.

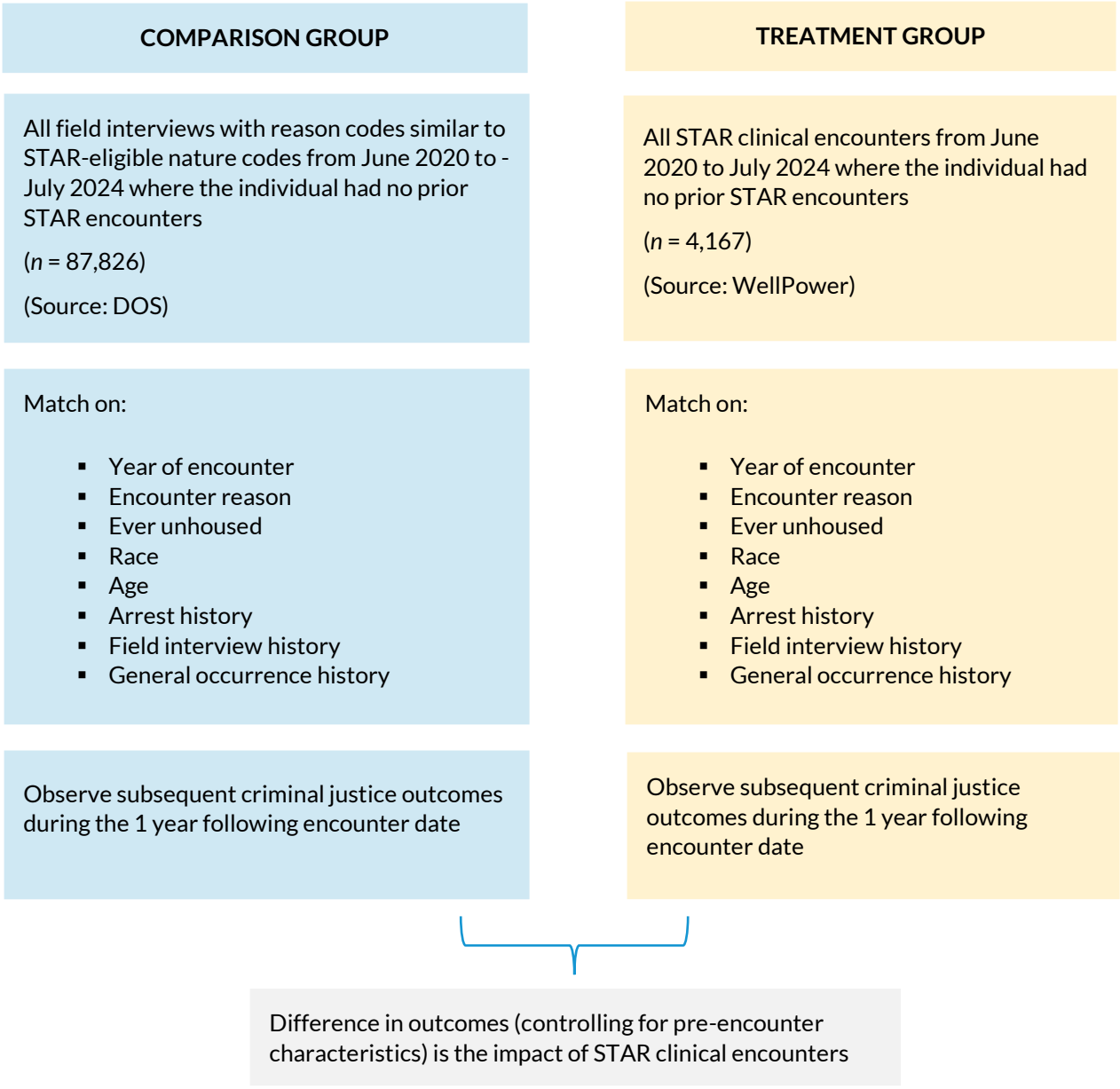
For each outcome, we measure its occurrence within one year after the focal encounter with STAR or the Co-Responder Program and report both continuous and binary measurements of that outcome. For example, for arrests, we look at the number of arrests within one year following the focal encounter (continuous measure), as well as whether there were any arrests within the same time horizon following the focal encounter (binary measures).

Despite a good balance between the treatment and comparison group, it is still prudent to include controls in the regression analysis. In each regression, we control for the same measures used to estimate propensity scores. We run ordinary least squares regressions for all outcomes using the matched sample with robust standard errors clustered at the person level.

Figures A1 and A2 illustrate the analysis methods described in this appendix and used to estimate the impact of STAR and co-responder clinical encounters, respectively.

FIGURE A1

Illustration of Analysis Method for STAR Van Clinical Encounters

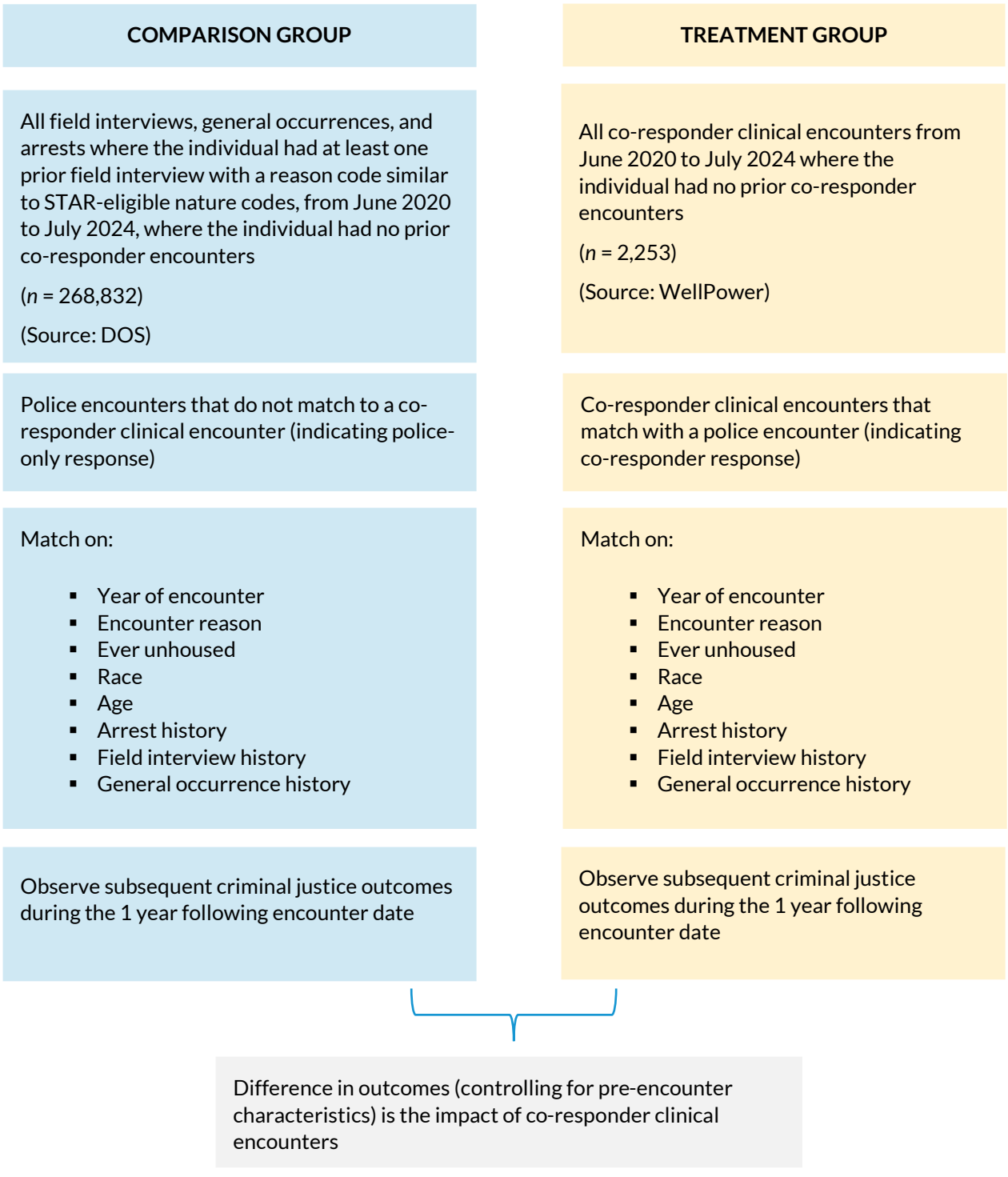


URBAN INSTITUTE

Source: Framework developed by the authors.

FIGURE A2

Illustration of Analysis Method for Co-Responder Clinical Encounters



Source: Framework developed by the authors.

Sensitivity Analysis

We perform a number of robustness checks to see if our results change significantly. First, we run unmatched regressions with the full comparison group, controlling for the same covariates as included in the propensity score matching process. We find the statistical significance and direction do not change; however, the magnitude is larger when using the unmatched data. Second, given the skewed distribution of many of our outcomes, we also run Poisson regression models. The statistical significance and the coefficient direction do not change from regressions run using ordinary least squares although the coefficients are generally larger in magnitude under the Poisson model. Next, we run pre/post models looking at just the treatment group. We compare the pre/post results between the treatment and comparison groups, and we see no change in statistical significance or direction. Lastly, we run a series of regressions where we subset our data to only include cases within a well-covered range of propensity scores because some treatment cases had very high propensity scores and there were few comparison cases with similarly high propensity scores. Again, these models have the same statistical significance and direction and very similar magnitudes as our full models. The models we present in this paper are likely the most conservative estimates of program impacts compared with other models used in our sensitivity analysis.

Appendix B. Regression Outcomes

TABLE B1.

Regression Outcomes—All STAR Clinical Encounters (Treated n = 4,167)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	0.41	0.48	0.07 ***	0.42	0.29	-0.12 ***	0.42	0.39	-0.03	0.42	0.39	-0.03
Police contacts	1.61	1.50	-0.11 *	1.62	1.44	-0.18 ***	1.62	1.46	-0.16 *	1.63	1.45	-0.18 ***
Bookings	0.28	0.33	0.05 ***	0.29	0.23	-0.05 ***	0.29	0.30	0.02	0.29	0.30	0
STAR encounters	0.05	0.46	0.42 ***	0.05	0.39	0.34 *** ^	0.05	0.36	0.31 *** ^	0.06	0.19	0.12 *** ^^^
Binary												
Arrests	0.17	0.19	0.01 *	0.17	0.16	-0.02 *** ^	0.18	0.15	-0.02 * ^^			
Police contacts	0.50	0.46	-0.04 ***	0.50	0.46	-0.04 *** ^	0.50	0.42	-0.08 *** ^^^			
Bookings	0.15	0.17	0.02 ***	0.15	0.15	0	0.15	0.15	0			
STAR encounters	0.03	0.22	0.2 ***	0.03	0.10	0.07 *** ^^^	0.03	0.13	0.1 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on STAR clinical encounters come from WellPower.
Notes: The STAR encounter sample is defined as first-time STAR clinical encounters between June 1, 2020, and July 30, 2024. The comparison encounter sample is defined as an eligible field interview (defined in Appendix A) over the same time period. The sample size is 4,167 in the STAR encounter group and 87,826 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. Statistical significance is denoted as: * = significant at 5%; **= significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = >.8.

TABLE B2.

Regression Outcomes—STAR Van Encounters

Subgroup: Encounters with clients whose race is documented as white (treated n =2,248)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	0.39	0.44	0.05	0.40	0.30	-0.1 *	0.39	0.36	-0.03	0.40	0.34	-0.05
Police contacts	1.53	1.45	-0.08	1.53	1.46	-0.07	1.54	1.41	-0.13	1.54	1.39	-0.15
Bookings	0.25	0.28	0.03	0.25	0.21	-0.04 *	0.25	0.27	0.01	0.26	0.25	0
STAR encounters	0.06	0.51	0.46 ***	0.06	0.43	0.37 *** ^	0.08	0.36	0.29 *** ^	0.09	0.20	0.12 *** ^^
Binary												
Arrests	0.16	0.18	0.03 ***	0.16	0.15	-0.01 ^	0.16	0.14	-0.01 ^			
Police contacts	0.48	0.48	-0.01	0.48	0.46	-0.02	0.48	0.43	-0.05 * ^^			
Bookings	0.13	0.15	0.02 *	0.13	0.13	0	0.14	0.13	0			
STAR encounters	0.03	0.23	0.21 ***	0.03	0.11	0.07 *** ^^^	0.04	0.12	0.08 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on STAR clinical encounters come from WellPower.

Notes: The STAR encounter sample is defined as first-time STAR clinical encounters between June 1, 2020, and July 30, 2024, with clients whose race is documented as white. The comparison encounter sample is defined as an eligible field interview (defined in Appendix A) over the same time period with clients whose race is documented as white. The sample size is 2,248 in the STAR encounter group and 39,355 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. * = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

TABLE B3.

Regression Outcomes—STAR Van Encounters

Subgroup: Encounters with clients whose race is documented as not white (treated n = 1,919)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	0.43	0.53	0.11 ***	0.44	0.30	-0.14 ***	0.43	0.42	-0.02	0.44	0.42	-0.01
Police contacts	1.68	1.55	-0.12	1.68	1.41	-0.27 ***	1.70	1.51	-0.18	1.70	1.50	-0.2 *
Bookings	0.31	0.38	0.07 ***	0.32	0.25	-0.07 *	0.31	0.35	0.03	0.32	0.34	0.02
STAR encounters	0.04	0.40	0.37 ***	0.04	0.34	0.3 *** ^	0.04	0.32	0.27 *** ^	0.05	0.16	0.1 *** ^^^
Binary												
Arrests	0.19	0.20	0.01	0.19	0.16	-0.03 *** ^^	0.19	0.16	-0.03 * ^^			
Police contacts	0.51	0.44	-0.07 ***	0.51	0.44	-0.08 *** ^^	0.51	0.42	-0.09 *** ^^^			
Bookings	0.16	0.18	0.02 *	0.16	0.16	-0.01	0.16	0.17	0			
STAR encounters	0.02	0.21	0.19 ***	0.02	0.09	0.07 *** ^^^	0.03	0.13	0.1 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on STAR clinical encounters come from WellPower.

Notes: The STAR encounter sample is defined as first-time STAR clinical encounters between June 1, 2020, and July 30, 2024, with clients whose race is documented as not white. The comparison encounter sample is defined as an eligible field interview (defined in Appendix A) over the same time period with clients whose race is documented as not white. The sample size is 1,919 in the STAR encounter group and 48,471 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. * = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

TABLE B4.

Regression Outcomes—STAR Van Encounters

Subgroup: Encounters with clients who have never experienced homelessness (treated n = 1,975)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	0.17	0.14	-0.03 *	0.17	0.18	0.01	0.17	0.17	0	0.18	0.16	-0.01
Police contacts	1.27	1.19	-0.09	1.27	1.32	0.06	1.27	1.30	0.03	1.27	1.27	0
Bookings	0.14	0.12	-0.02	0.14	0.15	0.02	0.14	0.14	0.01	0.14	0.14	0
STAR encounters	0.04	0.50	0.47 ***	0.04	0.44	0.4 *** ^	0.04	0.40	0.36 *** ^	0.04	0.15	0.11 *** ^^^
Binary												
Arrests	0.11	0.08	-0.02 ***	0.11	0.10	0	0.11	0.10	0			
Police contacts	0.45	0.41	-0.04 ***	0.45	0.44	-0.01	0.45	0.44	-0.01			
Bookings	0.09	0.07	-0.01 *	0.09	0.09	0.01 ^	0.09	0.09	0			
STAR encounters	0.02	0.23	0.21 ***	0.02	0.10	0.08 *** ^^^	0.02	0.11	0.09 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on STAR clinical encounters come from WellPower.

Notes: The STAR encounter sample is defined as first-time STAR clinical encounters between June 1, 2020, and July 30, 2024, with clients who have never been documented as experiencing homelessness. The comparison encounter sample is defined as an eligible field interview (defined in Appendix A) over the same time period with clients who have never been documented as experiencing homelessness. The sample size is 1,975 in the STAR encounter group and 74,166 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model.

TABLE B5.

Regression Outcomes—STAR Van Encounters

Subgroup: Encounters with clients who have experienced homelessness (treated n = 2,192)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	1.69	0.79	-0.9 ***	1.57	1.50	-0.07	1.56	1.46	-0.1	1.62	1.46	-0.16
Police contacts	3.46	1.78	-1.68 ***	3.30	2.83	-0.46 ***	3.28	2.84	-0.44 ***	3.33	2.68	-0.65 ***
Bookings	1.08	0.52	-0.56 ***	1.00	1.01	0.01	0.98	1.01	0.03	1.02	1.06	0.04
STAR encounters	0.11	0.43	0.32 ***	0.12	0.37	0.25 *** ^	0.12	0.34	0.22 *** ^	0.15	0.30	0.15 *** ^^
Binary												
Arrests	0.54	0.28	-0.26 ***	0.51	0.44	-0.08 *** ^^	0.52	0.44	-0.09 *** ^^^			
Police contacts	0.78	0.50	-0.27 ***	0.76	0.65	-0.11 *** ^^^	0.77	0.61	-0.16 *** ^^^			
Bookings	0.48	0.25	-0.23 ***	0.45	0.41	-0.04 * ^	0.45	0.47	0.02 ^			
STAR encounters	0.07	0.22	0.16 ***	0.07	0.16	0.09 *** ^^^	0.08	0.18	0.09 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on STAR clinical encounters come from WellPower.
Notes: The STAR encounter sample is defined as first-time STAR clinical encounters between June 1, 2020, and July 30, 2024, with clients who have ever been documented as experiencing homelessness. The comparison encounter sample is defined as an eligible field interview (defined in Appendix A) over the same time period with clients who have ever been documented as experiencing homelessness. The sample size is 2,192 in the STAR encounter group and 13,660 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. * = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

TABLE B6.

Regression Outcomes—STAR Van Encounters

Subgroup: Encounters with clients who are under 40 (treated n = 1,873)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	0.45	0.64	0.19 ***	0.47	0.29	-0.18 ***	0.46	0.40	-0.05	0.46	0.42	-0.04
Police contacts	1.64	1.65	0.01	1.65	1.40	-0.24 ***	1.66	1.44	-0.22 *	1.66	1.45	-0.21 *
Bookings	0.33	0.44	0.12 ***	0.33	0.23	-0.1 ***	0.33	0.31	-0.01	0.33	0.32	-0.01
STAR encounters	0.04	0.40	0.36 ***	0.04	0.32	0.28 *** ^	0.05	0.27	0.22 *** ^	0.05	0.13	0.07 *** ^^^
Binary												
Arrests	0.19	0.23	0.04 ***	0.19	0.17	-0.02 *** ^^	0.19	0.13	-0.06 *** ^^^			
Police contacts	0.50	0.47	-0.03 ***	0.50	0.44	-0.06 *** ^^	0.50	0.41	-0.09 *** ^^^			
Bookings	0.17	0.22	0.05 ***	0.17	0.16	-0.01	0.17	0.15	-0.02 ^^			
STAR encounters	0.02	0.20	0.18 ***	0.02	0.07	0.05 *** ^^^	0.03	0.09	0.06 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on STAR clinical encounters come from WellPower.
Notes: The STAR encounter sample is defined as first-time STAR clinical encounters between June 1, 2020, and July 30, 2024, with clients who are under 40 years of age. The comparison encounter sample is defined as an eligible field interview (defined in Appendix A) over the same time period with clients who are under 40 years of age. The sample size is 1,873 in the STAR encounter group and 50,782 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. * = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

TABLE B7.

Regression Outcomes—STAR Van Encounters

Subgroups: Encounters with clients who are 40 and over (treated n = 2,294)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	0.35	0.35	0	0.36	0.28	-0.07 *	0.36	0.32	-0.05	0.37	0.33	-0.04
Police contacts	1.57	1.38	-0.2 ***	1.57	1.46	-0.11	1.60	1.44	-0.16 *	1.60	1.43	-0.17 *
Bookings	0.22	0.23	0.01	0.23	0.21	-0.02	0.23	0.26	0.03	0.24	0.27	0.03
STAR encounters	0.06	0.52	0.46 ***	0.07	0.45	0.38 *** ^	0.07	0.40	0.33 *** ^	0.08	0.23	0.14 *** ^^^
Binary												
Arrests	0.15	0.15	0	0.15	0.15	0	0.15	0.14	-0.01 ^			
Police contacts	0.50	0.46	-0.04 ***	0.50	0.47	-0.03 *** ^	0.50	0.44	-0.07 *** ^^			
Bookings	0.12	0.13	0	0.12	0.13	0	0.13	0.15	0.02 ^^			
STAR encounters	0.03	0.25	0.22 ***	0.03	0.13	0.1 *** ^^^	0.04	0.14	0.1 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on STAR clinical encounters come from WellPower.

Notes: The STAR encounter sample is defined as first-time STAR clinical encounters between June 1, 2020, and July 30, 2024, with clients who are 40 years of age or older. The comparison encounter sample is defined as an eligible field interview (defined in Appendix A) over the same time period with clients who are 40 years of age or older. The sample size is 2,294 in the STAR encounter group and 37,044 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. * = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

TABLE B8.

Regression Outcomes—STAR Van Encounters

Subgroup: Encounters with clients who had one or more prior arrests (treated n = 906)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	1.31	1.52	0.21 *	1.32	1.29	-0.04	1.33	1.20	-0.13	1.34	1.24	-0.1
Police contacts	3.12	2.96	-0.16	3.13	2.85	-0.28	3.17	2.77	-0.41 *	3.17	2.76	-0.42 *
Bookings	0.87	0.97	0.1	0.88	0.90	0.02	0.89	0.89	0	0.90	0.88	-0.01
STAR encounters	0.07	0.47	0.4 ***	0.07	0.40	0.33 *** ^	0.09	0.38	0.29 *** ^	0.09	0.23	0.14 *** ^^^
Binary												
Arrests	0.47	0.50	0.03	0.47	0.45	-0.03 ^	0.48	0.43	-0.05 ^^			
Police contacts	0.75	0.72	-0.03	0.75	0.70	-0.05 *** ^^	0.75	0.68	-0.07 * ^^^			
Bookings	0.42	0.43	0.02	0.42	0.40	-0.02	0.42	0.40	-0.02			
STAR encounters	0.04	0.24	0.2 ***	0.05	0.12	0.07 *** ^^^	0.05	0.16	0.11 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on STAR clinical encounters come from WellPower.
Notes: The STAR encounter sample is defined as first-time STAR clinical encounters between June 1, 2020, and July 30, 2024, with clients who had one or more arrests prior to the focal encounter. The comparison encounter sample is defined as an eligible field interview (defined in Appendix A) over the same time period with clients who had one or more arrests prior to the focal encounter. The sample size is 906 in the STAR encounter group and 20,453 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. * = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

TABLE B9.

Regression Outcomes—STAR Van Encounters

Subgroup: Encounters with clients who had no prior arrests (n = 3,261)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	0.14	0.19	0.06 ***	0.14	0.07	-0.07 ***	0.13	0.14	0.01	0.14	0.13	-0.01
Police contacts	1.16	1.09	-0.06	1.16	1.04	-0.12 *	1.17	1.04	-0.12	1.17	1.03	-0.14 *
Bookings	0.10	0.15	0.04 ***	0.11	0.06	-0.04 ***	0.10	0.12	0.01	0.11	0.11	0
STAR encounters	0.04	0.46	0.42 ***	0.04	0.38	0.34 *** ^	0.05	0.32	0.27 *** ^	0.06	0.14	0.08 *** ^^^
Binary												
Arrests	0.08	0.10	0.02 ***	0.08	0.07	-0.02 *** ^^^	0.08	0.08	-0.01 ^			
Police contacts	0.42	0.39	-0.04 ***	0.43	0.38	-0.05 *** ^	0.42	0.37	-0.06 *** ^^			
Bookings	0.07	0.09	0.03 ***	0.07	0.07	0 ^	0.07	0.07	0			
STAR encounters	0.02	0.22	0.2 ***	0.02	0.09	0.07 *** ^^^	0.03	0.10	0.07 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on STAR clinical encounters come from WellPower.

Notes: The STAR encounter sample is defined as first-time STAR clinical encounters between June 1, 2020, and July 30, 2024, with clients who had no arrests prior to the focal encounter. The comparison encounter sample is defined as an eligible field interview (defined in Appendix A) over the same time period with clients who had no arrests prior to the focal encounter. The sample size is 3,261 in the STAR encounter group and 67,373 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. * = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

TABLE B10.

Regression Outcomes—All Co-Responder Clinical Encounters (Treated n = 2,253)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	1.06	0.55	-0.51 ***	1.06	0.96	-0.1 ***	1.06	0.96	-0.1 ***	1.06	0.89	-0.17 ***
Police contacts	3.32	2.02	-1.3 ***	3.31	3.18	-0.13	3.32	3.15	-0.16 *	3.32	2.93	-0.39 ***
Bookings	0.82	0.46	-0.36 ***	0.82	0.76	-0.05 *	0.82	0.75	-0.06 *	0.82	0.70	-0.12 ***
COR encounters	0.07	0.30	0.23 ***	0.07	0.26	0.19 *** ^	0.07	0.22	0.15 ***	0.07	0.15	0.07 *** ^^^
Binary												
Arrests	0.33	0.22	-0.11 ***	0.33	0.31	-0.01	0.33	0.28	-0.04 *** ^^			
Police contacts	0.67	0.57	-0.1 ***	0.67	0.67	0	0.67	0.65	-0.02 ^			
Bookings	0.35	0.25	-0.1 ***	0.35	0.36	0.01	0.35	0.34	-0.01			
COR encounters	0.03	0.17	0.14 ***	0.03	0.12	0.09 *** ^^^	0.03	0.15	0.11 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on co-responder clinical encounters come from WellPower.
Notes: The co-responder encounter sample is defined as an individual's first co-responder clinical encounters between June 1, 2020, and July 30, 2024. The comparison encounter sample is defined as any eligible field interview, general occurrence, or arrest (defined in Appendix A) over the same time period. The sample size is 2,253 in the co-responder encounter group and 268,832 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. * = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

TABLE B11.

Regression Outcomes—Co-Responder Encounters

Subgroup: Encounters with clients who have experienced homelessness (treated n = 750)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	2.66	1.21	-1.45 ***	5.11	4.72	-0.39 *	2.66	2.28	-0.37 ***	2.66	2.01	-0.66 ***
Police contacts	4.87	2.79	-2.08 ***	9.33	8.79	-0.54 *	4.86	4.30	-0.57 ***	4.87	3.96	-0.91 ***
Bookings	1.79	0.93	-0.86 ***	1.80	1.46	-0.34 ***	1.78	1.56	-0.22 ***	1.79	1.43	-0.35 ***
COR encounters	0.12	0.28	0.16 ***	0.18	0.33	0.15 ***	0.12	0.27	0.14 ***	0.12	0.25	0.13 *** ^^^
Binary												
Arrests	0.60	0.40	-0.2 ***	0.60	0.50	-0.1 *** ^^^	0.60	0.50	-0.09 *** ^^^			
Police contacts	0.78	0.64	-0.14 ***	0.78	0.71	-0.07 *** ^^^	0.78	0.73	-0.05 * ^^			
Bookings	0.60	0.45	-0.15 ***	0.60	0.54	-0.06 *** ^^	0.60	0.58	-0.03 ^			
COR encounters	0.05	0.17	0.12 ***	0.05	0.14	0.08 *** ^^^	0.06	0.18	0.13 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on co-responder clinical encounters come from WellPower.
Notes: The co-responder encounter sample is defined as an individual's first co-responder clinical encounters between June 1, 2020, and July 30, 2024, with clients who have ever been documented as transient in public safety data. The comparison encounter sample is defined as any eligible field interview, general occurrence, or arrest (defined in Appendix A) over the same time period with clients who have never been documented as transient in public safety data. The sample size is 750 in the co-responder encounter group and 85,725 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. * = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

TABLE B12.

Regression Outcomes—Co-Responder Encounters

Subgroup: Encounters with clients who have never experienced homelessness (treated n = 1,503)

	Unadjusted Means			Adjusted OLS/GLM			Adjusted PSM OLS/Logit			Adjusted PSM Poisson/Logit		
Events in the following year	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**	**Control**	**Treated**	**Diff.**
Continuous												
Arrests	0.45	0.22	-0.24 ***	0.45	0.41	-0.05 *	0.45	0.38	-0.07 ***	0.46	0.32	-0.14 *** ^
Police contacts	3.19	1.63	-1.56 ***	3.18	3.11	-0.07	3.18	3.04	-0.14	3.18	2.72	-0.47 ***
Bookings	0.42	0.23	-0.2 ***	0.42	0.40	-0.02	0.42	0.38	-0.04	0.43	0.33	-0.1 *** ^
COR encounters	0.08	0.31	0.23 ***	0.08	0.28	0.2 *** ^	0.08	0.22	0.13 ***	0.08	0.14	0.06 *** ^^
Binary												
Arrests	0.19	0.13	-0.06 ***	0.19	0.18	-0.01 ^	0.19	0.17	-0.02 ^^			
Police contacts	0.63	0.54	-0.09 ***	0.63	0.63	0	0.63	0.64	0.01			
Bookings	0.20	0.15	-0.05 ***	0.20	0.20	0	0.20	0.21	0			
COR encounters	0.03	0.17	0.14 ***	0.03	0.12	0.09 *** ^^^	0.03	0.13	0.1 *** ^^^			

Sources: Data on arrests, police contacts, and bookings come from the Denver Department of Public Safety. Data on co-responder clinical encounters come from WellPower.
Notes: The co-responder encounter sample is defined as an individual's first co-responder clinical encounters between June 1, 2020, and July 30, 2024, with clients who have never been documented as transient in public safety data. The comparison encounter sample is defined as any eligible field interview, general occurrence, or arrest (defined in Appendix A) over the same time period with clients who have never been documented as transient in public safety data. The sample size is 1,503 in the co-responder encounter group and 183,107 in the police encounter comparison group. The comparison group was identified via propensity score matching. "Adjusted" models included the following control measures: age, race, encounter reason, prior homelessness, and pre-enrollment measures of the criminal justice measures. "Unadjusted" models include no control measures. "Police contacts" include field interviews and general occurrences. "PSM" = propensity score matching. "OLS" = ordinary least squares regression. "GLM" = general linear model. For binary outcomes, logit regression models were used. * = significant at 5%; ** = significant at 1%; *** = significant at 0.1%. Effect sizes are denoted as: ^ = >.2; ^^ = >.5; ^^^ = > .8.

Notes

[1] Because of challenges staffing the program overnight, the program shifted to operating from 6:00 a.m. to 12:00 a.m. beginning in September 2025.

[2] Kirby Gaherty, "Behavioral Crisis Response in Minneapolis: Embracing Inclusive Design in Community Safety," National League of Cities, May 22, 2023.

[3] Denver 9-1-1 receives on average more than 100,000 emergency and nonemergency calls each month, but only generates about 45,000 calls-for-service each month in dispatched first responders. This is because many received calls pertain to the same incident (e.g., a crash on the highway that dozens of separate callers may report) and some calls do not require a response (e.g., a call for which an associated report could be filed online).

[4] We consider a vehicle to have "responded" if either (1) a vehicle (a STAR van or a police car, for example) notes that it has arrived at a scene or (2) the time between when a vehicle is assigned to a call for service and when that vehicle marks that call for service as cleared is greater than or equal to five minutes. We include this second condition because many records are missing the time when a vehicle arrived on scene, but for many of those calls, the amount of time between when a vehicle is assigned to and cleared from a call is quite substantial. We also take this approach because STAR vans may respond to a call for service via a phone call rather than physically driving to the location of a call for service and because a time of arrival is not applicable for a phone-based response. Due to this approach, call for service response estimates in this study may differ from those published elsewhere.

[5] During the 4.5 years covered in this study, there were 5,692 first-time STAR encounters, or 1,265 encounters annually. Multiplied by an estimated reduction in police contacts of -.16, this comes to 202 police contacts.

[6] All costs from our prior research are adjusted to 2023 dollars to align with the program costs used in this analysis.

[7] Denver 9-1-1 receives on average more than 100,000 emergency and nonemergency calls each month, but only about 45,000 calls-for-service each month result in dispatched first responders. This is because many received calls pertain to the same incident (e.g., a crash on the highway that dozens of separate callers may report) and some calls do not require a response (e.g., a call for which an associated report could be filed online).

[8] Callers may request a STAR-only response, indicating they only wish to have STAR respond and do not want a response from a police unit or other first responder agency. In some such cases, if STAR is not available to respond, no response will be dispatched.

References

Beck, Jackson, Melissa Reuland, and Leah Pope. T2020. “Shifting from Police to Community Responses. In: Behavioral Health Crisis Alternatives.” New York City: Vera Institute of Justice.

City and County of San Francisco Street Crisis Response Team. 2022. *Street Crisis Response Team Pilot: Final Report*. San Francisco: City and County of San Francisco.

Davis, Jonathan, Samuel Norris, Jadon Schmitt, Yotam Shem-Tov, and Chelsea Strickland. 2025. “Mobile Crisis Response Teams Support Better Policing: Evidence from CAHOOTS.” NBER Working Paper 33761. Cambridge, MA: National Bureau of Economic Research.

Dee, Thomas S., and Jaymes Pyne. 2022. "A Community Response Approach to Mental Health and Substance Abuse Crises Reduced Crime." *Science Advances* 8 (23). https://doi.org/10.1126/sciadv.abm2106.

Dee, Thomas S., and Jaymes Pyne. 2025. "Emergency Mental Health Co-responders Reduce Involuntary Psychiatric Detentions in the USA." *Nature Human Behaviour* (2025): 1-8. https://doi.org/10.1038/s41562-025-02339-7.

Gaherty, Kirby. 2023. Behavioral Crisis Response in Minneapolis: Embracing Inclusive Design in Community Safety. Washington, DC: National League of Cities.

Gillespie, Sarah, Devlin Hanson, Josh Leopold, and Alyse Oneto. 2021. Costs and Offsets of Providing Supportive Housing to Break the Homelessness-Jail Cycle: Findings from the Denver Supportive Housing Social Impact Bond Initiative. Washington, DC: Urban Institute.

Gillespie, Sarah, Mari McGilton, and Amy Rogin. 2023a. “Understanding Denver’s Co-Responder Program: Alternative Crisis Response in Denver.” Washington, DC: Urban Institute.

Gillespie, Sarah, Mari McGilton, and Amy Rogin. 2023b. “Understanding Denver’s STAR Program: Alternative Crisis Response in Denver.” Washington, DC: Urban Institute.

Gillespie, Sarah, Will Curran-Groome, and Amy Rogin. 2024. “Evaluating Alternative Crisis Response in Denver’s Support Team Assisted Response (STAR) Program: Interim Findings.” Washington, DC: Urban Institute.

Gillooly, Jessica, Tamara Leech, and Brenda Bond-Fortier. 2023. Transforming Denver’s First Response Model: Lessons in Multi-*Level* Systems Change. New York, NY: NYU School of Law Policing Project.

Gonzalez Miranda, Luis A., Anuradha Shetty, and Daniel Ehlke. 2024. “Analyzing Alternative Behavioral Crisis Response Models in the U.S.” *Journal of Community Health* 49: 324–29. https://doi.org/10.1007/s10900-023-01299-1.

Townley, Greg, and Emily Leickly. 2021. “Portland Street Response: Six-Month Evaluation.”. Portland, OR: Portland State University Homelessness Research & Action Collaborative.

Townley, Greg, Kaia Sand, and Thea Kindschuh. 2019. “Believe Our Stories & Listen: Portland Street Response Survey Report.” Portland, OR: Portland Street Response Community Outreach Workgroup.

Yamamoto, Marianna, Ana Obiora, and Gabriela Solia Torres. 2025. “Examining Alternative Response: A Landscape Analysis of Nine Community Responder Teams.” Boston: Harvard Kennedy School Government Performance Lab.

About the Authors

Sarah Gillespie is an associate vice president in the Housing and Communities Division at the Urban Institute, where her research focuses on solutions to homelessness. In addition to leading the evaluation of Denver's STAR and Co-Responder programs, she is a principal investigator for Denver's Housing to Health project, a seven-year experimental evaluation of a supportive housing pay-for-success initiative in Denver for people with unmet health needs and criminal justice involvement.

Will Curran-Groome is a research associate in the Housing and Communities Division. His work focuses on climate change adaptation, housing access and quality, land-use regulation, and active transportation.

Brendan Chen is a research analyst in the Housing and Communities Division. His research interests include housing affordability and homelessness.

Devlin Hanson is a principal research associate in the Family and Financial Well-Being Division. She is a labor economist whose research focuses on vulnerable children and families, including child welfare–involved families and immigrant families. She is a principal investigator for the Denver Housing to Health supportive housing evaluation, and she specializes in analyses using large longitudinal and cross-sectional administrative and public-use microdata.

Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org