Mobile Crisis Response Teams Support Better Policing: Evidence from CAHOOTS*

Jonathan Davis
University of Oregon

Samuel Norris
University of British Columbia

Jadon Schmitt
University of Oregon

Yotam Shem-Tov
UCLA and NBER

Chelsea Strickland
Texas A&M

February 19, 2026

Abstract

This paper studies the use of civilian-based mobile crisis response teams (CRTs)—a non-uniformed pair consisting of a mental health worker and a medic—as a component of emergency response to 911 calls. We provide the first evaluation of the longest-running CRT program in the United States, Crisis Assistance Helping Out on the Streets (CAHOOTS) in Eugene, Oregon, which responds to calls involving mental illness, homelessness, and addiction either instead of or in addition to police officers. We use two complementary research designs to understand the effects and possible scope of these programs. First, we find that a series of program expansions into new areas and times reduced the likelihood that a 911 call resulted in an arrest by 24 percentage points. The arrest reduction likely reflects the role of CRTs in de-escalating tense situations and resolving incidents without coercive measures. CRTs are most often dispatched to the same calls as the police, acting as an addition rather than a substitute. Second, we exploit idiosyncratic variation in CRT availability in the post-expansion periods to estimate the effect of additional marginal program expansions. We find that most marginal responses do not affect arrests and would otherwise go unanswered, suggesting that the program has reached a scale where it can respond to the most urgent calls.

*Authors can be reached at jdavis5@uoregon.edu, sam.norris@ubc.ca, schmittjadon@gmail.com, shem-tov@econ.ucla.edu, and cstrickland@tamu.edu, respectively. We thank Bocar Ba, Felipe Goncalves, Jens Ludwig, Grant McDermott, Charlie Rafkin, Marit Rehavi, Evan K. Rose, and seminar participants at multiple universities for their helpful comments. John Adegbonmire, Diti Jain and Everett Stamm provided outstanding research assistance. Funding was provided by Arnold Ventures and the Social Sciences and Humanities Research Council.

Police officers respond to a wide range of 911 calls, from violent crimes in progress to incidents involving mental health crises, homelessness, addiction, and interpersonal conflicts. Relying on police as universal first responders, however, may not be optimal: expanding officers' roles to include mediation and de facto social work can strain resources, reduce effectiveness in core law enforcement tasks, and increase the risk of unnecessary arrests or violence. An alternative approach is the integration of non-police first responders into municipal emergency-response systems. Typically composed of a non-uniformed mental health professional and a medic, these crisis response teams (CRTs) are trained to address crises related to substance use, mental health, and interpersonal conflict—situations that advocates argue differ fundamentally from traditional policing contexts and therefore warrant specialized responders (Irwin and Pearl, 2020; Krider et al., 2020).

Despite growing public demand for alternatives and complements to policing and the rapid expansion of these programs to more than 100 localities since 2020 (Ba et al., 2024, 2025), there is limited evidence on how CRTs affect the outcomes of 911 calls. Do they substitute for police responses, operate alongside police, or respond to calls that would otherwise go unanswered, thereby expanding the scope of emergency services? Importantly, do they successfully de-escalate crises and reduce arrests? And how do they affect broader measures of community safety? While there is an extensive literature on police officers, evidence on CRTs is sparse (Dee and Pyne, 2022; Ba et al., 2025).

This paper investigates the causal impact of sending CRTs to respond to 911 calls involving mental illness, homelessness, and addiction—either alone or as part of a broader police response. Our context is the longest-running such program in the United States, Crisis Assistance Helping Out on the Streets (CAHOOTS) in Eugene, Oregon. We use detailed administrative 911 call data and two complementary research designs to study the effect of the initial expansions and the effect of the marginal call after the program has reached scale, contributing to a comprehensive assessment of the effects and possible scope of crisis response programs.

CAHOOTS originated in Eugene, Oregon in 1989. At the start of our study period in 2014, it was active only in Eugene, running from mid-morning to late at night. Over the following four years, it expanded in five distinct steps to offer 24-hour coverage in Eugene as well as in the neighboring city of Springfield. Our first research design exploits this variation in a difference-in-differences framework, where we instrument for CRT responses to 911 calls based on the timing of the expansions. This strategy isolates the effect of CRT responses driven specifically by the introduction to new times and geographies. On average, the expansions increased the probability of a CRT response by 7.5 percentage points (pp), with a first-stage F-statistic of 173.

We find that CRT responses induced by the expansions reduced the likelihood that a call resulted in an arrest by 24 pp relative to a control complier mean of 32%. The reduction was driven by arrests authorized under an Oregon law that allows the detention of individuals who are a risk to themselves or others because of mental health issues or drug use. This is consistent with descriptive evidence suggesting that police responses to mental health crises increase the use of mental health holds (Walker et al., 2019), potentially because they have limited training to diagnose mental health conditions (Cohen and Bagwell, 2023). However, we see no impacts on arrests related to outstanding warrants, suggesting that the 911 dispatchers are appropriately targeting CRTs towards calls that require their expertise and unique training.

Some advocates of crisis response programs argue that they can serve as a substitute for the police in a wide variety of circumstances, with CAHOOTS itself arguing that they handle 17% of emergency calls and require police backup for only 1% of calls (White Bird Clinic, 2020). Conversely, critics argue that these programs mainly respond to calls that are outside the purview of the police and would not otherwise receive a response.

Our results suggest that neither of these views is correct, and that the truth lies somewhere between these poles. Approximately 40% of CRT responses induced by the expansions entail CAHOOTS responding alone, whereas counterfactually these calls would have received a police response. While this suggests that CRTs can sometimes substitute for the police, we also find that there are limits on the possible scope of substitution. In most of the remaining 60% of expansions-induced responses, the police arrive first to assess the situation for danger, after which a CRT is brought in and the police depart. Moreover, CAHOOTS responds to only 8% of 911 calls after the expansions. Thus, while CRT programs can potentially improve the quality of emergency response, they act as substitutes for the police only in limited situations.

Our second research design is tailored to identify the impacts of further expanding CRT availability. To achieve this, we introduce a novel "availability" instrument to assess a CRT's capacity at any given moment. The instrument is constructed by counting the number of active CRT calls when the focal call is received, and defining a call as high-availability if there is a below-median number of active calls. Through extensive validation exercises, we demonstrate that this measure can be treated as if it were randomly assigned within a specific city, month, day of the week and hour of the day (e.g., Wednesdays 9-10 p.m. in March 2024 in Eugene). The availability instrument strongly predicts whether a call receives a CRT response with an F-statistic of 523. These instrumented CRT responses lead to a moderate and statistically insignificant 4.5 pp reduction in the likelihood of an arrest, a striking contrast to the large effects we estimate in the expansions design.

To assess whether CRT responses have adverse effects on public safety, we examine impacts on several measures of criminal behavior. Under the expansions design, there are meaningful reductions in police crime reports—particularly for mental health, overdose, and drug-related incidents—with no statistically significant effects on reported violent or property crime. Because declines in reports may reflect changes in police reporting rather than underlying behavior, we also analyze future 911 calls from the same address, which is independent of a police response. CRT expansions reduce the likelihood of subsequent emergency and non-emergency calls in the following weeks, including declines in violent and serious calls, with no evidence of long-run increases. Availability-induced CRT responses are smaller and generally insignificant. Taken together, these findings suggest that CRTs do not compromise public safety and may reduce future demand for emergency services, particularly for higher-acuity incidents.

To better understand heterogeneity in the effects of CRTs, we develop a simple econometric framework that models the effect of a CRT response as arising through three possible channels: substitution (a CRT responds instead of the police), addition (a CRT responds to the same call as the police, in our case typically after the police have assessed the situation for danger), and service expansion (a CRT responds to calls that otherwise would have gone unanswered). We discuss conditions under which the complier means for each potential response can be identified, and find that more than 30% of CRT responses induced by the expansions would have resulted in an arrest had the police responded. In contrast, only 17% of the calls that were moved from a police-only response as a result of the availability instrument result in an arrest. Thus, while the composition of the responses explains some of the difference in effects across designs, we also demonstrate that the expansions generate responses to higher-acuity calls than additional availability, consistent with dispatchers appropriately targeting CRTs.

Through the lens of the model, effects can vary across designs through two channels: differences in effects (i.e., potential outcomes) and differences in the composition of complier types. Because the two designs operate on different margins—with the availability design affecting marginal responses after the program has grown to serve about 8% of calls—both channels may play a role in explaining the substantial differences in effects across designs. We find that approximately half of the CRT responses induced by the availability instrument would otherwise have received no emergency response. In contrast, nearly all responses induced by the expansion design would have been handled by the police in the absence of the program. However, differences in call composition explain only a minority of the heterogeneity in effects; most of the difference reflects variation in the effects of CRTs across designs. We view this pattern as most consistent with selection on gains, where calltakers successfully allocate calls based on the potential benefits of a CRT response, explaining the larger reductions in arrests for calls affected by the initial expansions than from those that receive a marginal CRT response after the program has reached scale.

We then turn to directly estimating the substitution and addition effects in the expansions design. Under the assumption that the complier means are the same across the five expansions—in other words, that differences in the reduced-form effects arise only because of differences in compliance patterns—we can identify the effects and potential outcomes for each of these two responses. We find large differences in baseline arrest risk: calls in which CRTs are added to a police response exhibit substantially higher counterfactual arrest risk than calls in which CRTs completely substitute for police. The effects of CRTs are correspondingly heterogeneous, and the arrest-reducing benefits of CRTs arise largely from coordinated responses with police, rather than from eliminating police involvement altogether.

Lastly, we combine the estimated benefits and costs of a CRT dispatch to calculate the marginal value of public funds (MVPF) under each research design. When dispatches are driven by an expansion in coverage, a CRT yields net cost savings, which implies an infinite MVPF so long as the willingness to pay is positive. For marginal expansions, the net cost is small—about $10— so the MVPF remains large even under conservative assumptions.

This paper contributes to several areas of research. First, a large literature studies the effects of police officers on crime and the optimal allocation of policing resources (Di Tella and Schargrodsky, 2004; Draca et al., 2011; Mello, 2019; Miller and Segal, 2019; Premkumar, 2019; Weisburst, 2019; Ba et al., 2021a,b; Chalfin et al., 2022; Hoekstra and Sloan, 2022; Mello, 2024; Rivera, 2025a,b). While increases in police staffing have been shown to reduce crime (Evans and Owens, 2007; Chalfin and McCrary, 2018), recent work highlights potential costs, including increased enforcement of low-level "quality-of-life" offenses, particularly for minority communities (Chalfin et al., 2022; Hoekstra and Sloan, 2022). In parallel, another strand of research has focused on improving police responses through training programs and technologies—such as Crisis Intervention Training (CIT) and body-worn cameras—and their effects on stops, arrests, and use of force (Owens et al., 2018; Rogers et al., 2019; Watson and Compton, 2019; Williams et al., 2021; Braga et al., 2022; Adger et al., 2025; Dube et al., 2025). Although this training-centered approach has long dominated policy responses to non-criminal incidents, evidence of its effectiveness has been mixed.

More recently, a growing literature has emphasized the role of non-police actors in improving public safety outcomes (e.g., Gonzalez and Komisarow (2020) study the effects of community monitoring of children's routes to school). We contribute to this debate by showing that expanding the set of emergency responders to include specialized civilian crisis response teams can substantially improve 911 outcomes by reducing arrests, likely by addressing immediate crises—such as mental health emergencies—that police are comparatively less well-equipped to manage. Importantly, these gains arise in settings where police departments have long operated CIT programs, suggesting that reallocating responsibilities across responders can be more effective than further investments in officer training alone.

However, our results also indicate that, as currently constituted, CRTs are not a full-fledged substitute for the police. CAHOOTS responds to only 8% of calls after the rollout, and does not respond to high-acuity calls. Additional availability resulting from lower-than-expected busyness goes towards responding to otherwise-unaddressed calls and does not have any effect on arrests. In this sense, CRT programs should be seen as having the potential to improve emergency response, rather than radically modify it.

Second, another body of research, primarily consisting of descriptive studies in medical and criminology journals, has examined co-response models, where a police officer and a civilian mental health professional jointly respond to incidents (e.g., Heslin et al., 2016; Jenkins et al., 2017; Keown et al., 2016; Blais et al., 2022). We contribute to this literature by providing a comprehensive evaluation of the causal effects of an all-civilian CRT program. This program is distinct from co-response models in that CAHOOTS staff do not respond in the same vehicles as police and, as we show, overlaps with the police officer for only about 50% of the time. Our program, CAHOOTS, is particularly important given the outsized role that it has played as a national template for police reform; cities including Denver, Minneapolis, Olympia, Rochester, and San Francisco (among others) have piloted their own versions.[1] As these programs continue to grow across the country, evidence of their effectiveness and interaction with existing emergency response services is key to their success.

Lastly, there are two recent studies most closely related to ours. Dee and Pyne (2022) study Denver's Support Team Assisted Response (STAR) program, which was modeled after CAHOOTS and launched in June 2020. Using a difference-in-differences design at the precinct–month level, they find that STAR reduced police reports of targeted offenses (e.g.,

[1]Often these programs are explicitly based on the CAHOOTS model (Fanelli, 2020). The model has also seen support from federal advisory committees (DOJ, 2015) and in proposed legislation (U.S. Senate, 2021); more recently the mayor of New York proposed that the city implement a similar program (Torres, 2025).

disorderly conduct, alcohol and drug use, and trespassing) by 34%. Ba et al. (2025) analyze the Holistic Empathetic Assistance Response Team (HEART) program, which began operating in Durham, North Carolina, in 2022. The HEART program includes multiple alternatives to a traditional police response: a civilian phone response, an unarmed in-person civilian response, and a co-response involving a mental health clinician and a CIT-trained police officer. They find that HEART responses reduced crime reports (57.5%), arrests (56%), and response times (11.8%).

We complement both studies in several ways. First, we use two complementary research designs to evaluate not only the effects of introducing CRT programs to new areas but also the potential gains from further expansions in program capacity. Second, we extend our 2SLS analysis by introducing a model that allows us to estimate the extent to which CRTs act as substitutes for the police or as additions to police responses, and understand the importance of each causal channel. Third, we examine a mature program—CAHOOTS has operated for over three decades—and provide evidence of its effectiveness largely prior to the COVID-19 pandemic. Finally, unlike aggregate crime report data, our detailed call-level data allow us to disentangle changes in outcomes from changes in reporting induced by reduced police presence.

2 Background and data

2.1 The crisis response program

The Crisis Assistance Helping Out on the Streets (CAHOOTS) program in Eugene, Oregon was established in 1989 by the White Bird Clinic to address crises involving mental illness, homelessness, addiction, and public assistance. As the acronym slyly suggests, CAHOOTS operates as a partnership with police and other emergency services in the area. Calls to 911 or the police non-emergency line are routed through the Lane County emergency communications center, which handles dispatch for police, fire, and ambulance, as well as CAHOOTS.

In contrast to co-response models, which involve teams of police officers and mental health workers who jointly respond to incidents, CAHOOTS utilizes a civilian-based approach for mobile crisis response. The CAHOOTS team consists of two people—a medic and a mental health crisis worker—in a Sprinter-style van loaded with emergency supplies. Medics may be a nurse, paramedic, or (most often) an EMT.[2] Crisis workers are not necessarily licensed

[2]Emergency Medical Technicians (EMTs) provide injury assessments, emergency medical care, and trans-

professionals, but typically have some experience in non-traditional mental health settings such as homeless outreach programs or crisis phone lines. Instead of police training, they receive specialized instruction on topics such as de-escalation, harm reduction, behavioral health, and substance abuse. Following classroom instruction, new team members complete 500 hours of field training, in which the trainee is supervised by two experienced team members. Upon successful completion of field training, the trainee begins working unsupervised in their own two-person team.

The team's goal during each call is to address callers' immediate needs, rather than to provide long-term solutions. This can include conflict resolution, suicide intervention, non-emergency medical care, connecting individuals to available programs such as drug rehabilitation support, and transporting callers to either a health care facility or a friend's house, as the situation warrants. As part of an effort to appear non-threatening, team members are unarmed and do not wear a uniform. They are also unable to make arrests, which they have cited as important for gaining the trust of the public.

2.2 The dispatching decision process

Calls to 911 in Eugene and Springfield are routed to the Lane County emergency communications center, where they are answered by specialized calltakers. The calltakers' job is to speak with the caller to understand the situation and record relevant information. As they do so, the computer-aided dispatch (CAD) system provides information on the address of the call and previous calls from the same number, as well as information about the caller from arrest and court records. Usually, within seconds of talking to the caller, the calltaker records the nature of the call from a pre-populated list of options; the most common call natures by response type are shown in Table 1.[3] Call natures are chosen by the calltaker during the call and directly influence the type of response that will be dispatched, or not. While calltakers also record extensive notes in the CAD system, we do not have access to this information.

Sitting in the same room as the calltakers are the dispatchers. The dispatcher's job is to communicate with emergency responders and assign specific units to calls. As the calltaker talks to the caller and adds information into the CAD system, the dispatcher decides which units—the specific police, fire, ambulance, or CAHOOTS vehicles—to assign to the call, and

portation to medical facilities. Compared to paramedics and nurses, EMTs are more limited in the type of care they can provide. For example, EMTs can deliver oxygen to patients, administer CPR, and provide Narcan for overdoses, but are not allowed to perform intubations or administer intravenous fluids.

[3] Figure A1 shows that Eugene and Springfield have a similar share of mental health related calls as other cities.

how to prioritize it relative to the currently outstanding calls. When call volumes are high, dispatchers leave low-priority calls in the queue and only assign a unit once there is sufficient availability. Some calls never get a response—for example, a welfare check where the caller phones back to say it is no longer required. The process is also dynamic. New information from the caller and from units that have already arrived on scene gets reflected in the CAD system, in the form of a change to the nature or updated notes. After the call has received a response or it has been deemed that a response is no longer necessary, the call is closed.

Unlike in other jurisdictions, dispatchers in our setting have the option of sending a CRT unit as part of the emergency response. While this sometimes entails dispatching a CRT unit as the sole responder, departmental guidelines restrict CAHOOTS from responding alone to calls that involve a crime or a potentially dangerous situation (Eugene Police, 2017). Dispatchers will therefore often first use a police unit to assess the level of danger, and then send a CRT unit if appropriate. In other situations, police that have arrived on scene will request a CRT unit. Thus, calls that involve both the police and CRT entail a combination of the different types of units working alongside each other on scene, and each type working alone.

2.3 Data sources and sample construction

Our analysis relies on computer-aided dispatch (CAD) data from the Eugene and Springfield police departments, which we link to arrest records for each city. Data for Eugene and Springfield includes all calls between January 1, 2014 and December 14, 2021, including both emergency and non-emergency calls.

We observe detailed information on each call including the date and time, nature (e.g., suicidal person, theft, robbery), priority, and location. Response information includes the time of each unit's dispatch and arrival, time of closing, and whether the call resulted in an arrest. We merge call data with census tract information. For each census tract, we observe a large number of characteristics including the size of the population, gender and racial makeup, median home values and rents, employment and unemployment rates, as well as per capita and household income.

For calls with an arrest, we also observe a summary of the charges used to justify the arrest. We use this information to separately look at arrests for involuntary mental health holds, detox, and warrants. Mental health holds require that an individual be identified as having a mental disorder and as being a threat to themselves or others (Gagnon et al., 2022); detox holds work similarly for intoxicated individuals.

Although our data include both 911 and non-emergency calls, our analysis of response patterns (i.e., whether a CRT or the police is dispatched) and outcomes focuses on 911 calls. We exclude non-emergency calls because they are less likely to result in a police dispatch.

2.4 Summary statistics

Table 1 reports information on 911 calls in Eugene and Springfield from January 1, 2014 to December 14, 2021. CRTs and the police are responsible for very different types of emergencies. Panel A reveals that CRTs disproportionately respond to calls for a wellness check and suicidal persons, but are not assigned to go alone to potentially violent situations such as robberies, fights, armed persons, or drunk drivers (Panel B). The police respond to a large number of traffic incidents (Panel C), as well as disorderly subjects, suspicious persons, and potential trespassing and burglaries. However, Panel D reveals that CRTs also respond to more potentially serious incidents alongside the police; 6.7% of disorderly subjects calls receive a response from both the police and a CRT.

Throughout the paper, we use nature codes as a coarse indicator of call characteristics. However, natures are assigned by calltakers based on judgment and may be updated as new information becomes available. While we observe each instance in which a nature is changed, we only have information about the final recorded nature. Panel E of Table 1 shows that more than 36,000 calls (13% of the total) have a changed nature; these calls are disproportionately police-only responses and exhibit above-average arrest rates. This pattern suggests that call natures may be endogenous to the response, and we therefore condition on natures only in robustness analyses.

These differences in types of calls reflect the differing specialties and training of CRTs and the police. CRTs are able to provide a range of on-scene services above and beyond those typically offered by the police, including first-aid and non-emergency medical care, suicide assessment and intervention, conflict resolution and mediation, and counseling for people grappling with grief and loss, substance use, homelessness, and other crises.

The differing acuity of calls across responder types is also reflected in the quickness of the response, and the time on scene. Panel A of Table 2 reveals that the average time to response is just over 64 minutes for a CRT-only call and about 24 minutes for a police-only call. This reflects the fact that CRT calls are often less urgent and so are slower to have a unit dispatched, as well as that there are fewer CRT vans available so they are slower to arrive once dispatched. Once on scene, a CRT stays for about 24 minutes for a CRT-only call and for about 41 minutes when they accompany the police. Police units are on scene for

52 minutes when responding alone and 72 minutes when responding with a CRT, although there may be multiple police units for some of that time.

Figure 1 examines 911 calls in which both police and a CRT respond, which—as shown in Table 1—tend to be higher-acuity cases. We decompose time on scene into periods with only a CRT present, only police present, and overlap between the two. The figure shows that approximately 50% of the time a CRT spends on scene overlaps with police presence (22 of 41 minutes). From the police perspective, overlap with a CRT accounts for about one-third of total time on scene (22 of 72 minutes), with the remaining time split roughly evenly between periods before a CRT arrives and after a CRT departs.

The figure also shows that CRTs typically arrive after the police, arriving second in 88% of joint calls. On average, a CRT spends 18 minutes on the scene after police departure, which constitutes the majority of its time on the incident. This pattern is consistent with the institutional structure described in Section 2.2: police often arrive first to assess scene safety, after which CRT units are dispatched. This sequencing is notably different from a co-response model, in which police and CRTs respond simultaneously and overlap for the entirety of the call.

The high rate of CRTs arriving after the police also assuages one concern with CRT programs: that they may expose unarmed first responders to potentially violent situations. To assess this concern more directly, we examine how police units are dispatched after a CRT unit arrives first on scene. Table A1 reports the results. Consistent with CAHOOTS' public statements—and the absence of publicized incidents involving injuries to CRT staff—no more than 3.5% of CRT responses involve a police unit being dispatched after CRT arrival, and fewer than 1% involve a follow-up police response that results in an arrest. These patterns suggest that police follow-up is rare and that the dispatch process generally succeeds in avoiding sending CRT teams alone to calls that are likely to escalate.

Panel B of Table 2 reveals that there are also differences in call outcomes across response types. Since CRTs are not authorized to make arrests, there are no arrests on CRT-only calls. In contrast, 14.5% of police-only calls and 13% of CRT-and-police calls result in an arrest, with much higher rates of non-criminal and detox holds among the CRT-and-police calls.

Finally, Panel C of Table 2 studies the reported outcomes of the calls. Reports can be filed only by the police, and so we see no reported outcomes on the CRT-only calls. Comparing police-only and CRT-and-police calls, however, we see a marked difference. While they have a similar likelihood of any report being filed, criminal activity is disproportionately represented among the police-only calls, while there are more mental health hold and detox reports among the calls with both types of units.

3 Effects of expansions in CRT services

3.1 Expansions research design

Our first research design seeks to quantify the impact of introducing CRT services on call outcomes. Specifically, at five distinct times during our study period, CRT vans suddenly became available for additional hours in either Eugene or Springfield.[4] Table A2 contains a list of these policy changes, which we use to estimate the effect of a CRT expansion in a difference-in-differences framework.

Figure 2 plots the number of 911 and non-emergency calls in each month for the relevant geography and hours of the day. For ease of interpretation, we normalize the number of calls by the number of hours in the design; for example, the first policy change in Eugene affects three hours (7-10 a.m.), so we divide the total number of calls each month by three. The changes in the number of calls are sharp and sudden, with the number of calls jumping at the time of expansion by 30-50 per hour-month in Eugene and 10-15 in Springfield.[5] Each of the changes are permanent, with the number of calls slightly increasing over the subsequent years. Focusing only on 911 calls, Figure A2 shows similar patterns. The number of CRT 911 calls per hour-month for each policy change jumps sharply and discontinuously at the timing of the expansions.

These graphs inform our first empirical strategy, which leverages the expansions as instrumental variables for a CRT dispatch. Specifically, we employ a simple difference-in-differences specification, which for a single expansion is structured as follows:

$$\begin{aligned} Y_i &= \beta \overbrace{C_i}^{\substack{\text{CRT}\\\text{response}}} + \psi^1_{t(i)} + \gamma^1_{c(i)} + u_i \\ C_i &= \alpha \underbrace{Z_i}_{\substack{\text{CRT}\\\text{expansion}}} + \underbrace{\psi^0_{t(i)} + \gamma^0_{c(i)}}_{\substack{\text{Calendar month}\\\text{\& city FEs}}} + \varepsilon_i \end{aligned} \tag{1}$$

[4] There were several policy changes in Eugene that added or removed a van, but without ever moving the number of available vans to zero. These policy changes had a much less dramatic effect on the likelihood that a call was answered by a CRT, so we focus on the extensive margin changes.

[5] Before each expansion, the number of calls that were dispatched to a CRT was very small but non-zero, and was the result of either vans working past the end of their shift to address additional calls, or dispatchers waiting for CRT vans to begin their shifts so they could assign a call to them.

where Y_i is an outcome such as a recorded arrest as a result of call i, C_i is an indicator for a CRT response, and Z_i is an indicator for the call coming in a time and geography following a CRT expansion (i.e., "Post $\times$ Treatment" in the standard difference-in-differences notation). We control for calendar month (e.g., May 2018) effects $\psi_{t(i)}$ and city effects $\gamma_{c(i)}$.

We observe five expansions and we therefore use a stacked design that combines all policy changes each as an instrumental variable for a CRT response. Since some of the designs overlap—for example, Table A2 reveals that the first Eugene expansion and the second Springfield expansion both included 9-10 a.m.—we additionally control for the other policy changes. Each design j consists of the hours of the day affected by the policy change, whether or not they are in the city where the policy change occurred. Specifically, we estimate the following 2SLS specification:

$$\begin{aligned} Y_{ij} &= \beta C_{ij} + \psi^1_{t(i)j} + \gamma^1_{c(i)j} + \sum_{j' \neq j} \pi_{jj'} Z_{ij'} + u_{ij} \\ C_{ij} &= \alpha_j Z_{ij} + \psi^0_{t(i)j} + \gamma^0_{c(i)j} + \sum_{j' \neq j} \alpha_{jj'} Z_{ij'} + \varepsilon_{ij} \end{aligned} \tag{2}$$

where Y_{ij} is an outcome on call i in design j and Z_{ij} is an indicator for the call coming in a time and city following the j^{th} expansion. $t(i)$ indexes months and $c(i)$ the two cities in our sample, so we control for design-month fixed effects $\psi_{t(i)j}$ and design-city fixed effects $\gamma_{c(i)j}$, analogous to time and location fixed effects in a traditional difference-in-differences specification. Some observations appear twice in this regression because of the design overlap, so we cluster standard errors at the call level. The stacked specification in Equation 2 combines all five expansions into a single over-identified 2SLS specification and can be interpreted as a weighted average of the IV estimates from each expansion separately.[6]

3.2 Balance checks

We begin by showing that there are no changes to the type or composition of 911 calls that coincide with the timing of the expansions (e.g., as a response to changing need for services). Figure 3 reports the coefficients from a regression of indicators for the call nature (i.e., calltaker description of the incident) on the expansions instrument. We find no effects on these call natures, nor on the likelihood of a police response, arrest, or an involuntary hold as

[6] The weights are proportional to the number of observations in the design multiplied by the residual variance in the instrument-predicted treatment. In our reported reduced form regressions, we use a single post-expansion indicator and variance-weight across designs to match the implicit design weights in the IV specification. This procedure ensures that our reduced form regressions are consistent with our main 2SLS specification and implies that if they are re-scaled by the first stage, they provide the same CRT effect estimate as our main 2SLS specification.

predicted by the call natures and geographic characteristics. Moreover, in robustness analysis below, we show that directly controlling for call natures does not impact our estimated effects.

Figure 3 also reports the coefficients from an OLS regression of the call natures and predicted outcomes on a CRT response. In contrast to the IV approach, there are large and precisely estimated correlations between a CRT response and the type of call. This highlights that simple comparisons between calls with and without a CRT response may be misleading due to confounding factors.

3.3 The effect of CRTs on call responses and arrests

Panel A of Table 3 reports reduced form estimates of the effect of an expansion on call outcomes. Consistent with Figure 2, the expansions caused a dramatic change in the response composition, with the likelihood of a CRT response jumping by 7.5 pp (F=173).[7] Simultaneously, column (2) shows that the likelihood of a police response decreased by 3.1 pp, suggesting that in about 40% of the instrument-induced responses a CRT acted as a replacement for the police. Column (3) reveals that there was nearly no effect on whether there was any response, implying that the remainder of the CRT responses—about 60%—were alongside the police.

To recast the reduced form estimates in terms of a CRT response, Panel B turns to estimating 2SLS regressions of the effect of a CRT response on call outcomes. We begin by describing the effects on the types of responders. Consistent with Panel A, column (2) of Panel B reveals that a CRT response reduces the likelihood of a police response by 41.4 pp relative to a control complier mean (CCM) of 96.2%, indicating that the expansion-induced CRT responses are for calls previously responded to by the police. However, as shown in column (3), there is nearly no effect on whether there is any response, with only 3.8% of marginal CRT responses counterfactually receiving no response.

The remainder of Panel B estimates the effect on call outcomes. A CRT response dramatically reduces the likelihood of arrest, by 24.3 pp relative to a CCM of 32.1%. This reduction is almost entirely explained by involuntary holds (15.1 pp) and detox holds (5.2 pp), suggesting that CRTs reduce arrests mostly by offering voluntary alternatives to non-criminal detention. We also find that CRTs have no effect on warrant arrests, consistent with the 911 center appropriately dispatching them to relatively low-risk calls. The CCM

[7]Before the expansions, a CRT responded to about 0.5% of calls. These responses occurred either when the caller phoned early in the morning and was placed on hold by the dispatcher until CRTs started work for the day, or when CRTs worked past their regular hours.

of arrest is more than three times the rate in non-CRT calls (see Panel C), indicating that expansion-induced CRT responses are directed toward calls where there is substantial scope to improve outcomes and where, in the absence of CRTs, coercive measures are more likely to be used.

Our analysis also reveals evidence of substantial omitted variable bias consistent with the imbalances documented in Figure 3, suggesting that simple comparisons of 911 calls with and without CRTs can be misleading. For instance, OLS estimates in Panel C suggest that a CRT increases involuntary hold arrests—an effect that is opposite in direction to the 2SLS estimates. Moreover, OLS underestimates the impact of CRTs on arrest, showing only a 3.7 pp reduction in arrests, compared to 2SLS estimates that show a substantively larger 24.3 pp reduction in arrests.

Lastly, Table A3 shows estimates of the effects of a CRT response, analogous to those in Table 3, for rare outcomes such as resisting arrest, use of force, either of these outcomes, or being dead on arrival. These outcomes are rare so we do not have sufficient precision to draw informative conclusions. We cannot reject meaningful positive or negative effect sizes.

The findings in Table 3 directly inform the question of the extent to which CRT programs can substitute for police responses. Under current dispatching criteria, CRTs function primarily as a complementary 911 service, providing additional tools for addressing emergency calls rather than replacing police involvement. Indeed, at the end of our study period, CRTs were dispatched to only 8% of calls. At the same time, our analysis indicates that calls receiving CRT responses have a high baseline risk, with a CCM arrest rate of 32.1%. CRTs substantially reduce arrests, suggesting that although they serve a relatively small share of 911 calls, they can generate substantial social welfare gains.

3.4 Robustness

We conduct a number of analyses to validate and check the robustness of the results in Table A4. First, we show our estimated effects are robust to the inclusion of additional controls. While our information on each call is limited, we observe the exact address, allowing us to map calls to their 2010 Census tract and measure a large number of tract characteristics from the ACS, including the size of the population, gender and racial makeup, median home values and rents, employment and unemployment rates, as well as per capita and household income. Panel A of Table A4 replicates our preferred specification. Panel B controls for these tract characteristics, and Panel C additionally adds the call nature as coded by the calltaker. Panel D instead includes tract-city-design fixed effects. Across each of these robustness exercises, we find remarkably similar results; for example, our estimate of the effect of a CRT response on arrests ranges only from -22.9 pp to -24.3 pp.

Second, in Table A5 we assess robustness by varying the sample based on the likelihood that a call receives a CRT response. Column 1 reports results for the full sample, as in Table 3. Column 2 restricts the sample to calls with a nature classification that ever receives a CRT response, yielding similar results. We then split calls by natures with high versus low CRT response rates.[8] Columns 3 and 4 report these estimates. The first stage is roughly four times larger for high-propensity natures, consistent with their greater suitability for CRT responses. The reduced form increases proportionally, yielding IV estimates that are similar and statistically indistinguishable across samples.

Third, to further validate the expansion instrument, we test whether placebo events affect arrest rates. Figure A3 displays the distribution of estimated effects for 500 randomly chosen event times, where the effects are estimated using the reduced-form specification to abstract from differing first stages across placebo times. The placebo event time effects are small, centered around zero, and do not overlap with the estimated effect from the observed data. These results are consistent with the estimates in Figure 3 showing that the CRT expansions are not correlated with changes in the nature of the calls or with tract characteristics.

Fourth, Table A6 presents reduced-form estimates corresponding to Equation 1 for each expansion separately. Reassuringly, all expansions increase the likelihood of a CRT response and decrease the likelihood of an arrest. Although the policy changes differ substantially in how much they raise the probability of a CRT response, the effects on arrests are similar for the Eugene and Springfield expansions. This consistency aligns with the over-identification J-test, which does not reject the null hypothesis that all expansions recover a common estimate of a CRT's effects on arrests ($p = 0.778$; see Table 3).

Fifth, we examine the robustness of our results to excluding the COVID-19 period. Specifically, Table A7 show that our main results are substantively unchanged if we restrict attention to calls before March 8, 2020, when Oregon declared a state of emergency because of the pandemic.

Sixth, one might worry that there were reductions in police capacity at the time of the CRT expansions, which might directly affect calls. Figure A4 shows that there were no appreciable changes in the number of units on duty at the time of the expansions. More directly, Table A9 controls for a measure of police availability (see Section 4) and finds almost

[8] As discussed in Section 2.2, observed call natures may partly reflect the type of unit dispatched, so these results should be interpreted with caution.

identical results.

Seventh, in Figures A5 and A6 we estimate the event study analogs to Equation 2, estimating the reduced form effect of the expansions on response type and arrest outcomes. Pre-trends are flat and the effects on treatment type and arrests are nearly constant over time, which makes the static difference-in-difference estimates a useful summary of the effects with more precision.

Eighth, we examine whether the expansions coincided with changes in the volume of 911 calls. Figure A7 plots the number of 911 calls per hour–month by city. As expected, given their relative size, Eugene receives substantially more calls than Springfield. However, we observe no break in call volume at the time of the expansions, consistent with the stable call composition documented in Figure 3. Table A10 further shows that our estimated effects are unchanged when we control for call volume.

Finally, we assess robustness to alternative ways of constructing counterfactual outcomes by estimating the reduced-form effect using both the traditional two-way fixed effects designs analogous to Equation 2 and the imputation approach of Borusyak et al. (2024). Since this approach uses only ever-treated city-hour-of-day cells, the sample is much smaller. Still, the results are very similar (see Table A11), further bolstering the robustness of our results.

4 Effects of marginal CRT responses

4.1 Availability research design

Using the expansions of CRTs to new times of the day and geographies provides an estimate of the effect of introducing CRTs as an additional option relative to a counterfactual where the police are the only available emergency service.[9] In this section, we study the likely effect of further additions to CRT capacity by estimating the effect of marginal CRT responses induced by quasi-exogeneous CRT availability *after* the initial expansions.

Our approach is to instrument for a CRT response with an indicator for there being a below-median number of active 911 calls to which CRT units are already assigned when the focal call is received. Specifically, for each focal call, we count the number of prior 911 calls that are still open separately for CRT- and police-dispatched calls, which—after residualizing out city-month-day of week-hour of day fixed effects to account for variation in available

[9] This is not strictly true, since there is a non-zero share of calls that are responded to by a CRT before the expansions. However, this occurs very rarely—in only 0.5% of calls—and only when callers wait for a CRT to become available or the CRT teams work overtime.

units and normalizing to have a mean of zero and a standard deviation of one—we call busyness (W_i). We then create a binary instrument $Z_i \equiv 1[W_i \leq \text{median}(W_i)]$ that indicates above-median CRT availability.[10] Because of the fixed effects, these comparisons focus, for instance, on 911 calls received in Springfield between 10 and 11 PM on Wednesdays in March 2021, comparing times with a high versus low level of CRT availability. We use a 2SLS specification, similar to Equation 1, in which a CRT response is instrumented by our measure of CRT availability:

$$\begin{aligned} Y_i &= \beta C_i + \gamma^1 W_i^p + \eta^1_{h(i)} + \varepsilon_i \\ C_i &= \alpha \underbrace{Z_i}_{\substack{\text{High CRT} \\ \text{availability}}} + \gamma^0 \underbrace{W_i^p}_{\substack{\text{Police} \\ \text{busyness}}} + \eta^0_{h(i)} + \varsigma_i \end{aligned} \tag{3}$$

where Y_i is an outcome such as a recorded arrest during call i, Z_i is the availability instrument, W_i^p is the number of open police calls (i.e., police busyness), and $\eta_{h(i)}$ are fixed effects at the city-month-day of week-hour of day level. Our design therefore compares calls received from the same area and within a very tight temporal window, but that were received at times of differing CRT availability.

4.2 First stage and balance checks

Panel A of Figure 4 visualizes the first stage effect of the availability instrument on the likelihood of a CRT response using the continuous measure of CRT availability. The relationship is negative and monotonic; an additional standard deviation of busyness (approximately 3.5 additional calls) decreases the likelihood of a CRT response by 2.1 pp (F-statistic = 442). A natural concern is that CRT availability is related to call characteristics even conditional on our rich set of fixed effects. To assuage this concern we conduct a battery of placebo tests where we regress observable call characteristics on CRT availability using the reduced-form analog to the IV specification.

The estimates in Figure 3 show that calls that come at times with more availability are not associated with differences in call natures. Each row reports the coefficient on call availability (blue solid square markers) or CRT assignment (green hollow circles) in a regression with the outcome given in the row header. CRT availability is uncorrelated with call natures and predicted response or outcomes. However, these characteristics are strongly predictive of actual assignment; as indicated by the hollow green circle markers, almost all of the

[10] Using a binary instrument makes the comparison to the expansions design results from Section 3 more straightforward and is consistent with the conceptual econometric framework that we discuss in Section 6.1.

characteristics are related to CRT assignment even conditional on the fixed effects. The predictive power of the call natures for actual assignment suggests that our placebo exercises should have some power to detect violations of exogeneity. This is particularly clear when the outcome is the likelihood of arrest as predicted using census tract characteristics and call natures. A CRT response is cross-sectionally associated with a 2 pp lower predicted arrest rate (mostly reflecting the difference between calls that do and do not receive any sort of response), but the call coming at a time of high availability has a negligible and statistically insignificant effect on the predicted arrest rate of 0.125 pp.

4.3 Effect of marginal CRT responses on call outcomes

Panel A of Table 4 reports reduced form estimates of the effect of high CRT availability on call outcomes. Consistent with Figure 4, the availability instrument leads to a sharp increase of 3 pp in the likelihood of a CRT response with an F-statistic of 523. Unlike the expansions instrument, marginal increases in CRT responses largely reflect answering calls that otherwise would have gone unanswered; higher CRT availability increases the likelihood of responding to a call by 1.5 pp.

Panel B of Table 4 turns to estimating 2SLS regressions of the effect of a CRT response on call outcomes. Column (2) reveals that a CRT response reduces the likelihood of a police response by 22.8 pp relative to a CCM of 50.3%, indicating that the availability-induced CRT responses crowd out a considerable amount of police response. This CCM is the likelihood of any response in the counterfactual; column (3) shows that an availability-induced CRT response increases the likelihood of any emergency response by 49.7 pp.

The remainder of Panel B estimates the effect on call outcomes. A CRT response decreases the likelihood of arrest by a statistically insignificant 4.5 pp. This coefficient is statistically different than the estimate derived using the expansion instrument, implying heterogeneity in effects across designs. While the standard errors are large enough that we cannot rule out modest effects, we also find small and statistically insignificant effects of a marginal CRT response on involuntary holds and warrant arrests. In contrast to the expansion results, we find a small but statistically significant 1.2 pp increase in detox holds.[11]

At a high level, we interpret these results as suggesting that further increases in CRT capacity yield smaller benefits than the initial expansions that extended coverage to new hours.

[11] Table A12 assesses the robustness of the estimated effects of CRTs to the inclusion of additional controls. As shown in the table, the results remain stable when controlling for census tract characteristics, call nature codes, and tract-by-city-by-design fixed effects, further reinforcing the credibility of the estimates. Table A8 also shows that the estimated effects are robust to the exclusion of the COVID-19 pandemic period.

Panel B of Figure 4 supports this interpretation: arrest rates are relatively flat—and not statistically distinguishable—across exogeneously-induced CRT response rates between 8% and 12%. We also present binscatter analogs to the reduced-form regressions in Figure A8, disaggregated by arrest type, which similarly indicate that additional capacity expansions have only modest effects on arrests.

Finally, Panel C of Table 4 examines the difference between the OLS and 2SLS estimates, and finds clear differences. For example, OLS estimates show a large increase in involuntary holds, contrary to the 2SLS estimates which show a null effect. We interpret this difference to mean that the OLS estimates are likely driven by omitted variable bias given the significant imbalances in call types documented in Figure 3.

5 CRTs and public safety

Our analysis thus far shows that dispatching CRTs can have substantial effects on call outcomes, with CRT responses induced by program expansions significantly reducing the likelihood of arrest. A potential concern, however, is that these reductions could be counterproductive in the longer run if these arrests would have deterred future criminal activity. Put differently, could CRT responses adversely affect public safety? To address this concern, we conduct several complementary analyses using a range of outcome measures.

In this section, we focus on reduced-form estimates rather than 2SLS effects, as our objective is to assess broader policy effects rather than the effect on a focal call. However, these estimates can be rescaled by the first stages to provide an effect that is interpretable in per-response terms.

5.1 Effects on crime reporting

Crime reports are the most commonly used measure of criminal behavior. Our data includes police reports that list suspected offenses associated with each call, whether or not an arrest takes place. Similar reports in other cities are often used to create measures of aggregate crime, making this a potentially comparable measure of the broader effects of CRTs.

Table A13 reports reduced-form estimates of the effects of the instruments on a wide range of crime-report outcomes. The expansions led to a statistically significant 1.2 pp reduction in the likelihood of any crime report, relative to a baseline rate of 22.5% in the pre-expansion period (Panel A). The largest declines are observed for mental health-related reports, which fall by 1.1 pp (a 58% reduction relative to baseline), and overdose-related reports, which decline by 0.3 pp, or approximately 40%. We also see some evidence of a decline in reports of criminal activity, with drug-related offense reports declining by a significant 0.244 pp, nearly half of the baseline rate of 0.6%. Effects on violent and property crime reports are negative but less precise and not statistically significant. By contrast, under the availability design (Panel B), estimated effects are generally smaller and statistically insignificant across most categories, with the exception of a small positive effect on drug-related reports that is marginally significant at the 10% level.

Taken at face value, this analysis suggests that a CRT does not have adverse effects and may even improve public safety. However, one important caveat is that because reports are filed exclusively by police officers, these effects could partially reflect changes in reporting rather than changes in crime. In particular, if the instruments replace police officers who would have filed a report with a CRT team who would not, reported crime would mechanically decline even if the underlying behavior remained unchanged. This concern is most relevant for call types in which a CRT and police are close substitutes, which is consistent with the large drops in reports of mental health and detox holds. By contrast, there should be relatively little reporting bias for more serious crimes, since CRT is not allowed to be dispatched alone to calls where there is a risk of danger. We therefore view this report-based analysis as informative for reported violent crime, and suggestive evidence that CRTs have not had a harmful effect on public safety.

5.2 Effect on future calls from the same address

Next, we analyze the effects of the program on the likelihood of a future 911 call from the same location, which we interpret as reflecting either an unresolved underlying issue or subsequent criminal activity. This outcome circumvents the concern that CRTs may mechanically reduce reported crime by substituting for police responses, since the measure does not rely on police-generated reports.

For each of the 90 days after the time of the focal call, we measure whether there was another 911 call to the same address within that time, excluding the first 24 hours to avoid capturing anything related to the focal call itself. We then use this as an outcome in reduced-form regressions on the expansions and availability instruments, allowing us to study the dynamics of the effects on future calls. Figure 5 reports the results. The red line in Panel A shows that after the CRT expansion, the likelihood of a future 911 call from the same location declines over the medium term. For the first several weeks, the likelihood of another call from the same address declines by about 2 pp, relative to the pre-expansion means of about 25% (Figure A9). The longer-run point estimates indicate some permanent reduction in future calls; however, the effects are not statistically significant.

The blue line in Panel A of Figure 5 reports the impact of the availability instrument on future 911 calls from the same address. The point estimates are negative for the four weeks following the focal call; however, they are statistically insignificant and—consistent with the smaller effect on arrests—smaller in magnitude than the effects in the expansions design. In the longer term, the point estimates are close to zero and are statistically insignificant. We interpret this as evidence that, even at the margin, CRT responses appear to reduce the need for future emergency services in the short run and have no detectable adverse long-run effects.

In Panel B we examine the effects on future non-emergency calls. The expansions cause a sharp, immediate, and persistent decline in the likelihood of a future call, while the availability instrument has nearly no effect. Since non-emergency calls are more likely to include calls that individuals make for themselves,[12] we view this as a complementary indicator of how the CRT responses caused by the expansions can reduce the need for future emergency response services.

We also evaluate the effects on more granular sets of calls. To do so, we define indicators for whether a 911 call with a violent or serious nature occurs within t days. We define serious calls to include all violent incidents as well as several additional natures with a high arrest rate and police response rate.[13] We then estimate the effect of a CRT response on receiving each of these types of calls.

Panels C and D of Figure 5 report these results. We find steep reductions in both violent and serious calls in the weeks following the focal incident, with magnitudes roughly half as large as the overall decrease in 911 calls. We take this as evidence that a CRT response can reduce not only nuisance calls but potentially more serious offenses. However, we find no such effects for the availability instrument, again consistent with the availability-induced responses largely going to otherwise-unanswered, lower-acuity calls.

We conduct several robustness analyses using alternative measures of future calls. Figure A10, Panel A measures future calls using anything received within 25 meters of the focal

[12] In fact, calls to the non-emergency number listed on CAHOOTS' website are actually routed to the Lane County communications center and answered by calltakers.

[13] Violent calls make up 22% of all calls, and include all those with a nature containing any of the following strings: shots fired, elderly abuse, armed, assault, bomb threat, child abuse, fight, gunshot wound, homicide, kidnap, riot, robbery, stab wound, weapon offense, sex abuse, hit and run injury, rape, officer involved shooting, harassment, dispute, dispute family. Serious calls are 27% of calls and include violent calls as well as burglaries, hit and runs, and any other nature with an arrest rate higher than the 85th percentile and police response rate higher than 70%.

call, while Panel B shows the estimated impact on future calls within 25 meters of the focal call, excluding the exact location. In both cases, we see declines in the likelihood of a future call. Finally, Panel C excludes generic addresses such as intersections and entire blocks, as well as locations with more than 300 911 calls; we also see declines in the likelihood of a future call from these addresses where displacement may be less likely.

6 Heterogeneity in the effects of CRTs: causal channels and program scope

Our analysis has shown that CRT responses can operate in two distinct ways. First, CRTs can substitute entirely for a police response. Second, CRTs can respond to incidents that still require police dispatch; in these cases, police typically arrive first to assess risk, followed by a period of overlap before the police depart to leave a CRT on scene alone.

This distinction might help explain an empirical pattern documented in this paper: CRT responses induced by higher-than-usual availability lead to a modest and statistically insignificant reduction in arrests (Section 4), whereas CRT responses induced by program expansions reduce arrests by nearly six times as much (Section 3). A natural explanation is treatment effect heterogeneity driven by differences in counterfactual responses. The objective of this section is to formalize these mechanisms and estimate heterogeneity in the effects of CRT responses across these scenarios.

6.1 An econometric framework for interpreting effects

We split the responses to a 911 call into four possible options: no response (n), only a CRT (c), only the police (p), and both a CRT and the police (b). Let $D_i \in \mathcal{D} \equiv \{n, c, p, b\}$ denote the response to call i, and let $Y_i(d)$ denote the potential outcome of interest under treatment d.[14]

To tractably describe the different counterfactuals we will focus on the case of a binary instrument $Z_i \in \mathcal{Z} \equiv \{0, 1\}$ that increases the likelihood of a CRT response. For a given instrument, there are therefore $|\mathcal{D}|^{|\mathcal{Z}|} = 16$ potential *response types*, or types defined by the set of counterfactual treatments for each instrument value (Heckman and Pinto, 2018).

[14] This treatment definition abstracts from some potentially important features, such as which type of unit arrives first in a b call. In Appendix A1, we use an additional research design to explore this avenue in more detail and find limited evidence that the order of arrival has a substantive effect on call outcomes, which we view as supportive of a more parsimonious treatment definition.

Table 5 presents the set of potential response types under each research design. Guided by our institutional setting, we assume that the instrument does not affect the likelihood of a police response except through substitution. For example, Z_i may move calls from p to c, but not from p to n.[15] Formally, this is stated as:

Assumption 1 (No cross effects except substitution)**.**

$$
\begin{aligned}
\mathbb{1}[D_i(1) \in \{b\}] &\geq \mathbb{1}[D_i(0) \in \{b\}] \ , \\
\mathbb{1}[D_i(1) \in \{n, c\}] &\geq \mathbb{1}[D_i(0) \in \{n, c\}] \ , \\
\mathbb{1}[D_i(1) \in \{c, p, b\}] &\geq \mathbb{1}[D_i(0) \in \{c, p, b\}] \ .
\end{aligned}
$$

Panel A of Table 5 shows how this assumption is sufficient to restrict the compliance patterns in the expansions design such that the instrument affects treatment for only three groups: calls that are moved from p to b, calls that are moved from p to c, and calls that are moved from n to c. Each of these response types therefore faces a different change in their treatment: for the $p \rightarrow b$ compliers a CRT is an *addition* to the police, for the $p \rightarrow c$ compliers a CRT is a *substitute* to the police, and for the $n \rightarrow c$ compliers there is a *service increase* since their calls would otherwise not receive a response. One implication of this assumption is that the estimated size of each response group should be positive; we show below that this is true.

In the availability design, we require an additional monotonicity assumption that the instruments only increase the likelihood of response involving CRT:

Assumption 2 (Monotonicity)**.**

$$
\mathbb{1}[D_i(1) \in \{c, b\}] \geq \mathbb{1}[D_i(0) \in \{c, b\}]
$$

This distinction is due to the fact that because there were no CRT units available before the expansions, Assumption 2 is satisfied by construction in the expansions design. Panel B of Table 5 shows how this additional assumption rules out the additional possible compliance groups in the availability design.

One potential channel through which these assumptions could be violated is if the instruments affected police dispatch behavior—for example, by freeing up police time and inducing officers to respond to additional calls. Although the positive mass on each compliance group

[15] We can relax this assumption to be conditional on a measure of the busyness of the police. We use a measure of police busyness similar to our CRT availability instrument as an additional control in robustness analysis below.

suggests this is unlikely to be a first-order concern, Table A9 shows that our main results are robust to directly controlling for police busyness. Moreover, Figure A11 shows that variation in police busyness is unrelated to the likelihood of a CRT response, consistent with CRTs being dispatched to calls for which they are appropriate responders rather than as substitutes for police during busy periods. This pattern is consistent with the absence of $b \to c$ compliers. Taken together, these pieces of evidence support the plausibility of our assumptions in this setting.

6.2 Decomposing and characterizing 2SLS estimates

Given Assumptions 1 and 2 , the Wald estimator of the effect of a CRT response can be decomposed as:

$$\frac{E[Y \mid Z{=}1] - E[Y \mid Z{=}0]}{E[\underbrace{1[D \in \{c,b\}]}_{\substack{\text{CRT}\\ \text{response}}} \mid Z{=}1] - E[1[D \in \{c,b\}] \mid Z{=}0]} = \underbrace{\frac{\pi_{pb}}{\pi_{pb}+\pi_{pc}+\pi_{nc}}}_{\equiv\, s_{pb}} \underbrace{\overbrace{E[Y(b) - Y(p) \mid p \to b]}^{\equiv \Delta_{p\to b}}}_{\substack{\text{Adding CRT to}\\ \text{a police response}}} \quad (4)$$
$$+ \underbrace{\frac{\pi_{pc}}{\pi_{pb}+\pi_{pc}+\pi_{nc}}}_{\equiv\, s_{pc}} \underbrace{\overbrace{E[Y(c) - Y(p) \mid p \to c]}^{\equiv \Delta_{p\to c}}}_{\substack{\text{Substituting from}\\ \text{police to CRT}}} + \underbrace{\frac{\pi_{nc}}{\pi_{pb}+\pi_{pc}+\pi_{nc}}}_{\equiv\, s_{nc}} \underbrace{\overbrace{E[Y(c) - Y(n) \mid n \to c]}^{\equiv \Delta_{n\to c}}}_{\substack{\text{CRT response}\\ \text{instead of no response}}}$$

where we denote compliers of type $D_i(0){=}d$, $D_i(1){=}d'$ as $d \to d'$ and where the shares of the complier groups, $\pi_{dd'} \equiv P[D(0){=}d, D(1){=}d']$, can be identified from changes in observed treatments:

$$\begin{aligned}
\pi_{pb} &= E\big[1[D_i{=}b] \mid Z_i{=}1\big] - E\big[1[D_i{=}b] \mid Z_i{=}0\big], \\
\pi_{pc} &= E\big[1[D_i{\in}\{n,c\}] \mid Z_i{=}1\big] - E\big[1[D_i{\in}\{n,c\}] \mid Z_i{=}0\big], \\
\pi_{nc} &= E\big[1[D_i{\in}\{c,p,b\}] \mid Z_i{=}1\big] - E\big[1[D_i{\in}\{c,p,b\}] \mid Z_i{=}0\big].
\end{aligned} \quad (5)$$

Equation 4 highlights that variation in 2SLS estimates across instruments can arise from differences in the treatment effects Δ, or from differences in the complier shares s. As a first step towards characterizing possible differences in treatment between the availability and expansion designs, Figure 6 reports the shares of compliers from each response type, $s_{dd'} = \pi_{dd'}/(\pi_{pb} + \pi_{pc} + \pi_{nc})$. We find sharp differences between the designs. Under the expansions design, CRT responses occur in combination with a police dispatch ($p \to b$, which includes responses alongside or after the police) in 54.8% of calls and in place of police dispatch ($p \to c$) in 41.4% of calls, while service expansion ($n \to c$) accounts for only 3.8% of cases. By contrast, under the availability design, 49.7% of induced CRT responses correspond to calls that would otherwise receive no response, with substantially fewer cases involving either addition or substitution.[16]

We view these results as suggesting that there are a relatively limited number of 911 calls where a CRT can operate as a complete substitute for the police, and after the initial expansions—which boosted the share of CRT calls from approximately 0 to 8%—most of these calls have been exhausted. Instead, most of the availability-induced CRT responses are to calls that otherwise would not have received a response. Further expansions to CRT capacity—by adding additional vans during the day, for example—would likely match this pattern.

Our potential outcomes framework also allows us to characterize the responses of the $p{\to}b$ compliers. In particular, we examine how police officers and CRTs interact when responding to the same 911 call. This analysis parallels Figure 1 but focuses on the compliers rather than all b calls; the estimates are broadly similar. Across both designs, about 50% of CRT time on scene overlaps with police presence, with most CRT-only time occurring after police departure. Police-only time differs across designs: it is roughly evenly split before and after CRT presence in the expansions design, but concentrated after CRT departure in the availability design. Overall, the patterns in Figure A13 confirm that joint police-CRT responses are not well characterized as approximating co-response models. Rather, a b response typically involves police arriving first to assess the situation and subsequently bringing in a CRT when appropriate, in contrast to co-response models in which police and CRTs work together throughout the incident.

6.3 Cross-design differences in complier potential outcomes and selection on gains

Understanding the causes of the difference in effects across the expansions and availability designs is key to understanding the likely effect of other possible policy changes. To shed

[16] As discussed above, one testable implication of Assumption 1 is that each of the estimated π's should be positive; Figure 6 shows that this condition is satisfied. Our assumption also implies that these shares should be positive in each subsample. In Figure A12, we estimate them separately for the periods of day affected by each of the expansions; we cannot reject that the estimated πs are non-negative in both the expansions (Panel A) and availability (Panel B) designs.

light on the potential outcomes underlying the effects, in Appendix A2 we show that:

$$\frac{E\big[Y1[D{=}b] \mid Z{=}1\big] - E\big[Y1[D{=}b] \mid Z{=}0\big]}{E\big[1[D{=}b] \mid Z{=}1\big] - E\big[1[D{=}b] \mid Z{=}0\big]} = E[Y(b) \mid p{\to}b] \tag{6}$$

$$\frac{E\big[Y1[D{=}p] \mid Z{=}1\big] - E\big[Y1[D{=}p] \mid Z{=}0\big]}{E\big[1[D{=}p] \mid Z{=}1\big] - E\big[1[D{=}p] \mid Z{=}0\big]} = \frac{s_{pb}}{s_{pb}{+}s_{pc}} E[Y(p) \mid p{\to}b] + \frac{s_{pc}}{s_{pb}{+}s_{pc}} E[Y(p) \mid p{\to}c] \tag{7}$$

where s are the known complier shares.[17]

This equation reveals that while the data are directly informative about potential outcomes under b for $p{\to}b$ compliers, they contain information on only a weighted average of $Y(p)$ for $p{\to}b$ and $p{\to}c$ compliers, rather than directly identifying each response group-specific $E[Y(p) \mid g]$. These average potential outcomes, however, are informative for understanding the differences in results across designs.

Table 6 contains estimates of the moments in Equations 6 and 7 for both the expansions and availability designs. Panel A reveals that $E[Y(b) \mid p{\to}b]$ is nearly identical across designs; if both the police and a CRT respond as a result of either instrument, the likelihood of arrest is 13-14%. In contrast, the average value of $Y(p)$ for $p{\to}b$ and $p{\to}c$ compliers varies wildly across the designs, from 0.321 in the expansions design to 0.159 in the availability design (Panel B). This difference cannot be explained by the weights s on the complier means; the table also reveals that they are almost identical across designs.

As a result of these differences in potential outcomes, the effects of CRTs vary dramatically across designs, even accounting for cross-design differences in the share of $n{\to}c$ compliers. We provide direct evidence on the magnitude of this gap by estimating the effect of CRTs among $p{\to}b$ and $p{\to}c$ compliers using the 2SLS analog of the following Wald estimator:

$$\frac{E[Y \mid Z{=}1] - E[Y \mid Z{=}0]}{E[1[D \neq p] \mid Z{=}1] - E[1[D \neq p] \mid Z{=}0]} = \underbrace{\frac{s_{pb}}{s_{pc} + s_{pb}} \Delta_{p\to b} + \frac{s_{pc}}{s_{pc} + s_{pb}} \Delta_{p\to c}}_{\text{Rescaled effect}} . \tag{8}$$

Panel C of Table 6 contains the results. Consistent with the cross-design differences in mean potential outcomes, we estimate very different effects across the availability and expansion designs (-8.9 pp and -24.6 pp, respectively).

Overall, we view this as strong evidence that the marginal call changed dramatically from the

[17] The likelihood of arrest is zero for $D \in \{n, c\}$, which implies that $E[Y(c) \mid p{\to}c] = E[Y(c) \mid n{\to}c] = E[Y(n) \mid n{\to}c] = 0$. In Appendix A2, we show that this means that the identifying variation in the instruments can be simplified from eight moments reflecting the Wald estimator of the effect of the relevant treatment on $Y1[D{=}d]$ or $Y1[D{\neq}d]$ to only the two in Equations 6 and 7.

time of the expansions—when CRTs responded to nearly no calls—to the post period, when CRTs had reached scale. The CRT responses induced by the expansion had a very high risk of arrest in the counterfactual where the police responded. After the expansions, however, the remaining marginal calls—which are captured by the availability instrument—were less pressing, and so had a substantially lower risk of arrest.

6.4 Disentangling causal channels

Since the previous analysis does not separately identify $E[Y(p) \mid p \to b]$ and $E[Y(p) \mid p \to c]$, it does not provide any direct evidence on the relative magnitude of the addition effect $\Delta_{p \to b}$ and the substitution effect $\Delta_{p \to c}$. One way to view this is that the treatment effects are simply underidentified: a single binary instrument provides only two moments to identify three parameters.

To overcome this difficulty, we note that one could generate the moments in Equations 6 and 7 for each of the five expansions. Then, under the assumption that the complier means are the same for each expansion, one can directly estimate these complier means using a minimum distance procedure to fit the system of ten linear equations in three unknowns. The plausibility of this approach is buttressed by Table 6, which shows that we cannot reject the null hypothesis that $E[Y(b) \mid p \to b]$ is the same across the five expansions, consistent with constant complier means.

Table 7 contains the results. Panel A reveals that the average arrest rates with a police-only response differ dramatically across the $p \to b$ and $p \to c$ groups: nearly 50% of the former group would be arrested by the police while only 12% of the latter group would be.[18] We view this difference as reflecting the varying severity of calls across these response groups; the calls that receive a CRT in addition to the police are of much higher risk than the calls where they substitute for the police.

These differences in complier means correspond to large heterogeneity in the effects of CRTs. Panel B of Table 7 reports these effects; substituting a police response with a CRT response reduces the arrest rate by 12 pp (SE=18.4 pp) while adding CRTs to a police response reduces the arrest rate by 35.7 pp (SE=17.4 pp). We conclude that much of the arrest-reducing benefit of a CRT comes from working together with the police—in our context, this usually means arriving at a call after the police, overlapping with them, and carrying on after they leave—rather than as an alternative.[19] These findings suggest that CRTs can

[18] Note that the 0.141 estimate for $E[Y(b) \mid p \to b]$ is slightly different than the 0.142 estimate in Table 6; this is because the latter is estimated via 2SLS rather than variance-weighted minimum distance.

[19] Panel C of Table 7 reports the results of additional over-identification tests, both overall and for the

function as effective complements to the police—an important dimension that has received little attention in scholarly and policy discourse.

Another way to illustrate the variation exploited by the minimum distance estimator is to use a visual IV-style figure (Holzer et al., 1988; Angrist, 1990) that plots reduced-form effects on arrests against reduced-form effects on different response types. Panel A of Figure A14 plots, for each expansion, the effects on arrests and on the likelihood of any CRT response, revealing substantial variation across expansions in both arrest reductions and CRT response rates. Panels B and C show similarly large variation across expansions in the effects on arrests, CRT-only responses, and responses involving both police and a CRT—precisely the variation that identifies the minimum distance estimates. Moreover, consistent with the results in Table 7, Figure A14 shows that variation in responses involving both police and a CRT is much more predictive of arrest reductions than variation in CRT-only responses: the R^2 is roughly three times larger, and the slope coefficient is about four times as large.

7 Welfare and cost trade-offs

CRTs can generate substantial cost savings when they replace police responses. However, because many CRT responses occur alongside police or address calls that would otherwise receive no response, the net fiscal impact of a crisis responder program is ambiguous. In this section, we use a simple marginal value of public funds (MVPF; Finkelstein and Hendren, 2020) framework to quantify net costs and assess how large program benefits must be for CRTs to be cost-effective at their observed scale in Eugene and Springfield, as well as for marginal expansions.

Table 8 summarizes the inputs to our MVPF calculation. In 2022, CAHOOTS' contract with the City of Eugene totaled \$759,854 (in 2021 dollars). Based on the most recent year in our sample, we estimate that CAHOOTS responded to approximately 2,993 calls annually, implying an average cost of \$253.88 per 911 response.[20] We assume constant costs, so the cost of an average and marginal response are equal. This assumption is conservative: if marginal costs are lower than average costs, we are underestimating the MVPF.

A key feature of the MVPF framework is that program costs should be adjusted for fiscal

$E[Y(p)]$ moment. For both tests we find p-values above 0.93, consistent with our assumption of constant complier means across expansions.

[20]We do not count non-emergency responses in formulating the cost-per-response, which we view as conservative in the sense of putting no value on these responses. It also reflects the discrete nature of the program; since for most of the study period Eugene and Springfield each have a single van on the streets at any time it is not possible to reduce quantities.

externalities. Our results indicate two primary sources of fiscal externalities: substitution between CRT and police responses and changes in arrests. The magnitude of these externalities differs for average versus marginal responses, reflecting differences in estimated treatment effects across designs.

We estimate the fiscal externality from substitution by scaling the cost of a police response by our estimated effect of a CRT response on the probability of police involvement, separately for the expansion and availability designs. The Eugene Police Department estimates that an average police response costs $931. Each call in which a CRT replaces police therefore saves $931, on average. Applying our estimated substitution effects implies a fiscal externality of –$385 per average CRT response and –$212 per marginal response.

We use an analogous approach to estimate fiscal externalities from reduced arrests. We assume each arrest costs $707, based on the estimated cost of a mental health hold in the Pacific Northwest (Karaca and Moore, 2020). This assumption is appropriate in our setting as the majority of arrest reductions are for mental health–related incidents. This implies a fiscal externality of –$172 per average response and -$32 per marginal response.

Combining these components, the net fiscal cost of an average CRT response is –$303, while the net cost of a marginal response is approximately $10. Thus, CAHOOTS is cost-saving on average and close to budget-neutral at the margin.

Table A13 shows that CRTs may lead to broader crime reductions, although the estimates are imprecise. If we take these estimates at face value, they suggest a further fiscal externality from crime reduction of -$392 for an average response and of -$46 for a marginal response and reductions in victim costs of $23 and $3. Here, we follow the approach of Deshpande and Mueller-Smith (2022) to map our estimates into changes in enforcement, court, and victim costs, conservatively using cost estimates for assaults, drug possessions, and thefts for our violent, drug, and property estimates.

When an intervention is not cost-saving, computing the MVPF requires assumptions about the value of social benefits in order to estimate willingness to pay (WTP). That step is unnecessary for average responses: because CAHOOTS is cost-saving, the MVPF is infinite so long as WTP is positive. For comparison, Ba et al. (2025) estimate an individual WTP of $69 per year (in 2021 dollars) for the related HEART program using contingent valuation adjusted for social desirability bias.

For marginal expansions, the MVPF depends on benefit valuation, but net costs are sufficiently small that the MVPF remains large even under conservative assumptions. For example, if each adult in Eugene is willing to pay just $1 per year for CAHOOTS, the ratio of total WTP to net costs is approximately 5; if WTP equals \$69, this ratio exceeds 350. Moreover, if the marginal cost of a response is even modestly lower than the average cost, the net marginal cost becomes negative and the MVPF is again infinite.

Overall, we show that the MVPF of CAHOOTS is infinite or very large without relying on strong assumptions about the value of program benefits. Cost-effectiveness is driven primarily by fiscal externalities from reduced police involvement and fewer mental health–related arrests, rather than by assumptions about difficult-to-measure social benefits. Incorporating fiscal externalities and social benefits from crime reductions further amplifies the MVPF estimates.

The MVPF does not account for several potentially important but difficult-to-monetize benefits of CAHOOTS. First, CAHOOTS predominantly responds to calls in poorer neighborhoods, making it a progressive public service. Figure A15 shows that census tracts with a higher share of CAHOOTS responses tend to have lower average rents (Panel A) and higher shares of non-White residents (Panel B), indicating that—beyond reducing arrests—the program disproportionately benefits low-income and marginalized populations. Second, the MVPF analysis excludes other hard-to-value benefits, such as improvements in emergency response times. As shown in Figure A16, CRT expansions reduce response times under both research designs. While difficult to monetize, faster response times reduce wait times for other callers and may also improve clearance rates (Blanes i Vidal and Kirchmaier, 2018).

8 Conclusion

We provide new evidence on the effectiveness of CAHOOTS, a crisis response program in Eugene and Springfield, Oregon that has served as a model for cities across the country grappling with emergencies involving homelessness, mental health crises, and addiction. Using rich 911 call data, we find that crisis response teams decrease the likelihood of arrest on the focal call—regardless of whether they respond instead of or in addition to police. CRTs also appear to effectively solve problems over the medium-term, decreasing the likelihood of a subsequent call from the same address over the next several weeks.

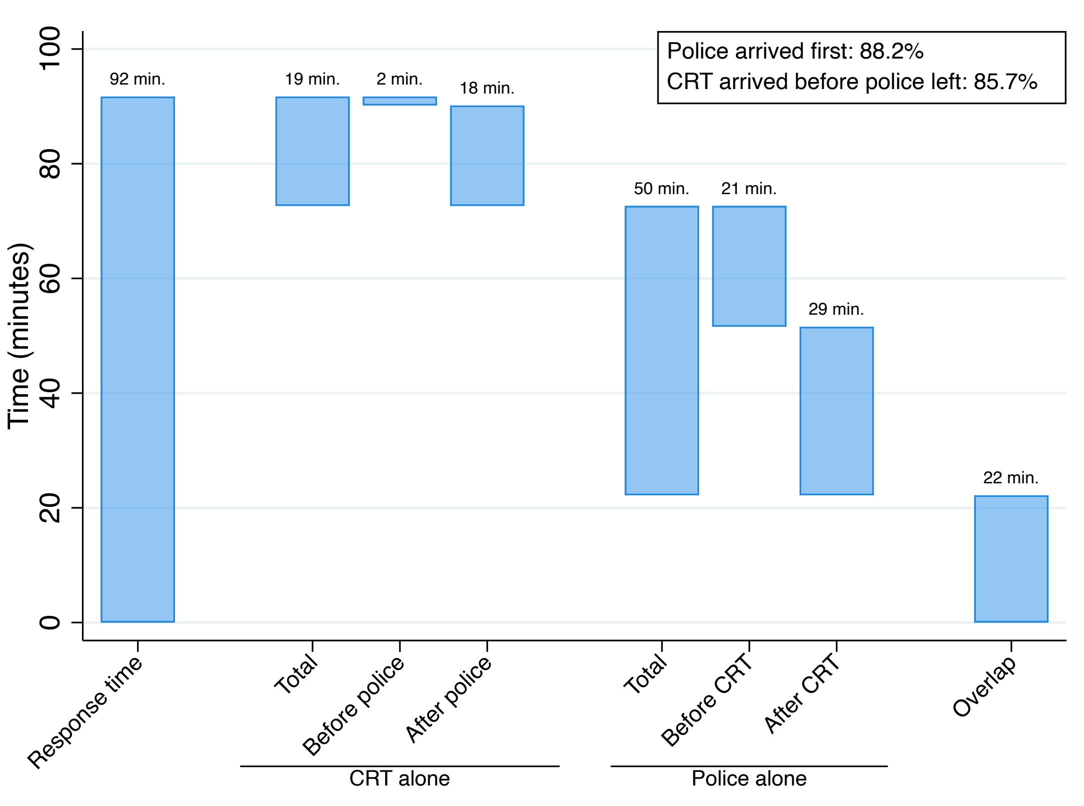
The arrest reductions associated with CRT responses are larger than those documented for many recent public safety reforms. For example, Tebes and Fagan (2025) find that ending stop-and-frisk reduced stop-related arrests by 61.1%. In contrast, under the expansions design, we estimate that a CRT response reduces the likelihood of arrest by 75.7%. Relatedly, Ba et al. (2021b) study officer demographic shifts in Chicago and find that, per 100 shifts worked, Black officers make 1.93 fewer arrests than White officers–approximately 21% of the average arrest volume for White officers. The corresponding effects are smaller for Hispanic officers (5%) and female officers (7%). That our estimated effects are larger is perhaps unsurprising, as CRTs fundamentally alter the occupation, training, skills, and legal authority of the responder, rather than changing only the demographic characteristics of police officers.

However, the ability of CRTs to serve as a substitute for the police is limited. We find that in the initial expansions that made CRTs suddenly available in new times of day or geographic areas, about 41% of CRT responses were substitutes for the police (in the other 59%, CRTs responded alongside the police). However, after the initial expansions, when CRT responds to about 8.3% of calls, marginal responses prompted by idiosyncratic availability of CRT units are mostly to calls that otherwise would have received no response. Future research should determine the extent to which these patterns are caused by statutory restrictions on the types of calls that CAHOOTS can respond to, or whether there are more general limits to the crisis responder model. This need is particularly acute given the ongoing evolution of crisis response across the country and suggests a need for research that pinpoints the characteristics of crisis response programs that are most predictive of effectiveness.

Our results are informative about the potential benefits and optimal scale of CRTs for the many localities considering the adoption of similar programs—nearly 100 new CRT programs have started since 2020 (Ba et al., 2025). In fact, Eugene, OR is now among the cities considering how to best implement a new CRT after the unexpected collapse of CAHOOTS in April 2025 when the White Bird Clinic said it did not have the financial resources to continue operating in the city. Lane County introduced a new countywide program (Lane County Mobile Crisis Service). Even with this program, the popularity of CAHOOTS has led the city to release a request for proposals for a new city CRT program with the results still pending.

Figures

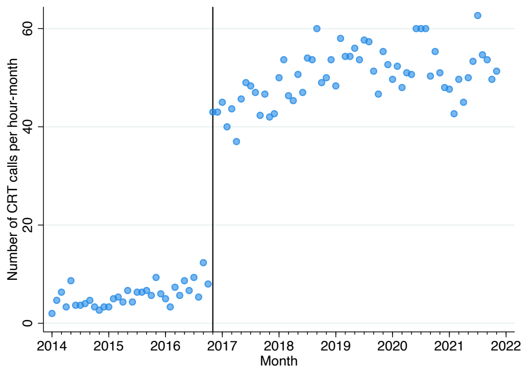
Figure 1: Characterizing responses with both police and a CRT



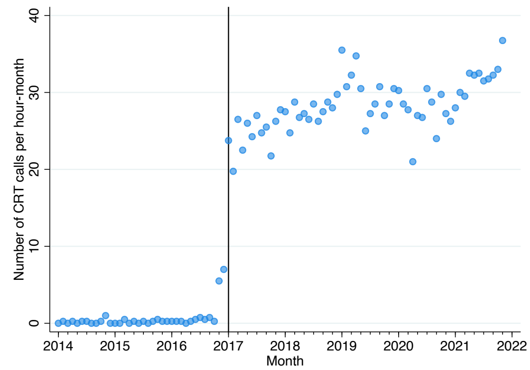
Notes: This figure reports the average number of minutes with each type of response among calls with both a police and CRT response. Overlap refers to minutes when both a CRT and police were on scene; the minutes alone for each group are broken into time before and after the other unit arrived.

Figure 2: CRT expansions in Eugene and Springfield

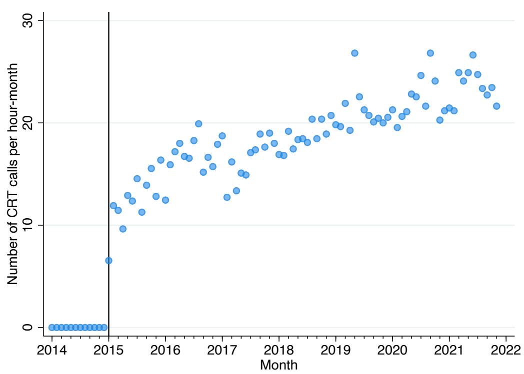
(a) Eugene 7-10 a.m.



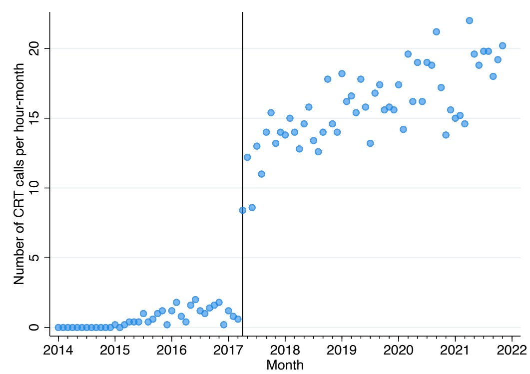
(b) Eugene 3-7 a.m.



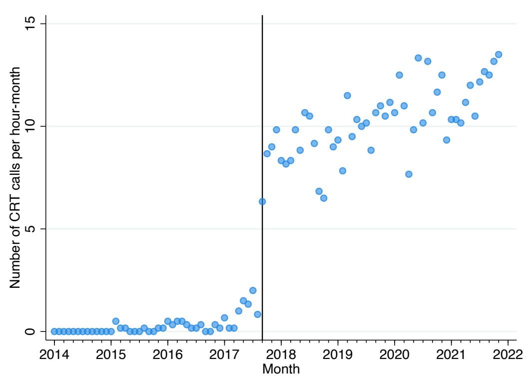
(c) Springfield 12 p.m. to 11 p.m.



(d) Springfield 9-11 a.m. and 11 p.m.-2 a.m.

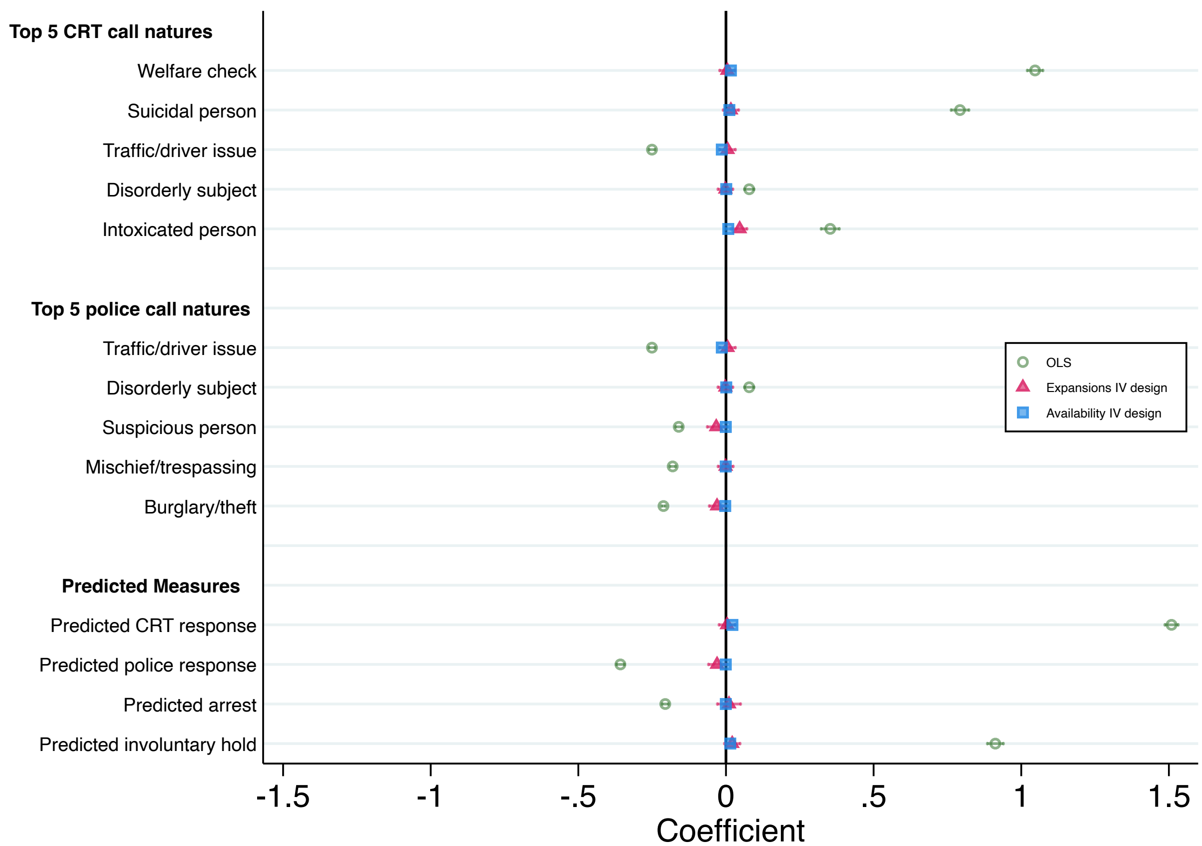


(e) Springfield 3-10 a.m.



Notes: This figure shows the changes in the number of calls to which a CRT is dispatched as services are expanded (Eugene) or introduced (Springfield). Panels A and B report expansions in Eugene and Panels C, D, and E expansions in Springfield. Each point in the figures reports the number of calls (911 and non-911) to which a CRT was dispatched, normalized to be measured in terms of calls per hour-month. Before each expansion, the number of calls that were dispatched to a CRT was very small but non-zero, and were the result of either vans working past the end of their shift to address additional calls, or dispatchers waiting for CRT vans to begin their shifts so they could assign a call to them.

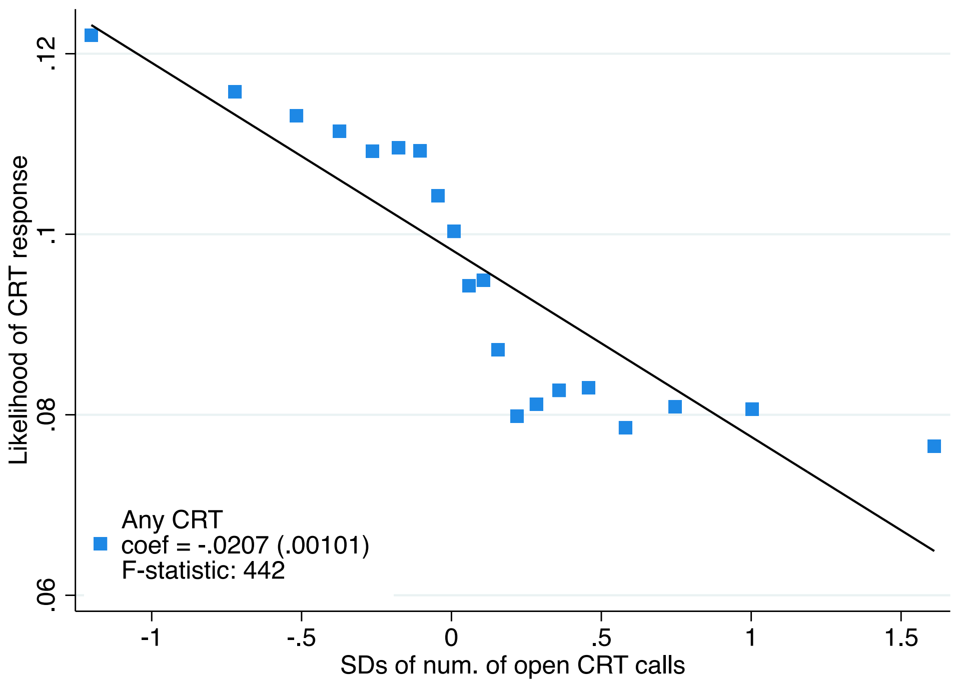
Figure 3: Balance checks of CRT expansions and availability instruments



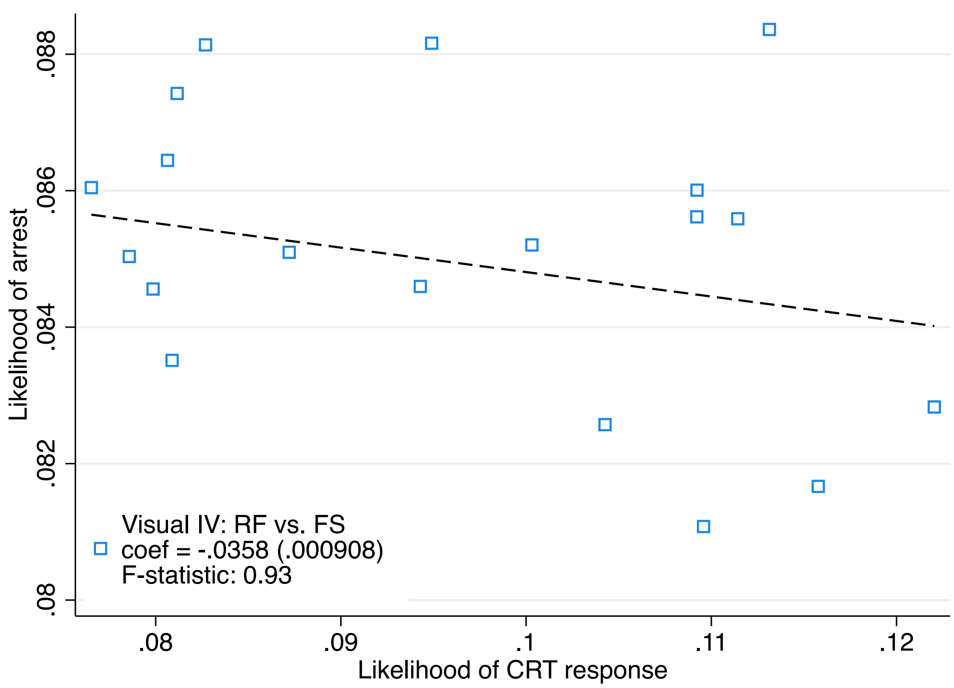
Notes: The figure reports estimates of the association between a CRT response and call natures as well as predicted measures based on natures and the geographic characteristics of the call location (green hollow circle markers). It also includes the association between CRT expansions and call characteristics (red solid triangular markers), and between the binary CRT availability instrument and call characteristics (blue solid square markers). All outcome variables have been standardized to have a mean of zero and a standard deviation of one to be on a comparable scale for the figure.

Figure 4: Busyness and the likelihood of a CRT response and arrest

(a) First stage of CRT response on busyness



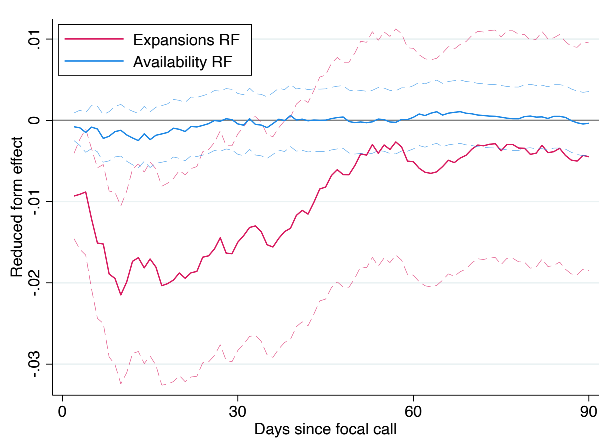
(b) Visual IV of arrest likelihood on instrument-induced CRT likelihood



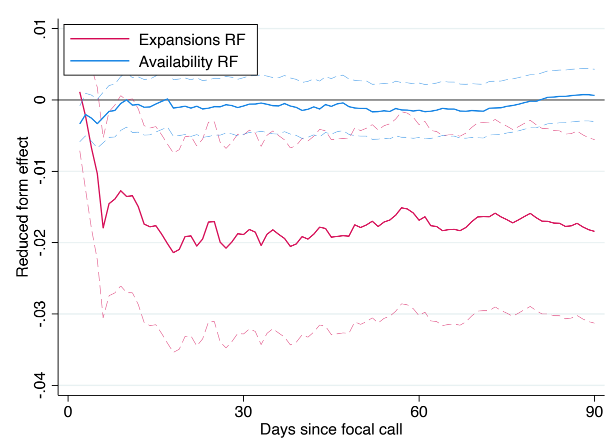
Notes: Panel A shows the first-stage relationship between the standardized busyness measure for a CRT and the likelihood of getting a CRT response. Panel B reports a visual instrumental variable style figure of the first stage relationship against the arrest rate.

Figure 5: Reduced form effect on future 911 and non-emergency calls from same address

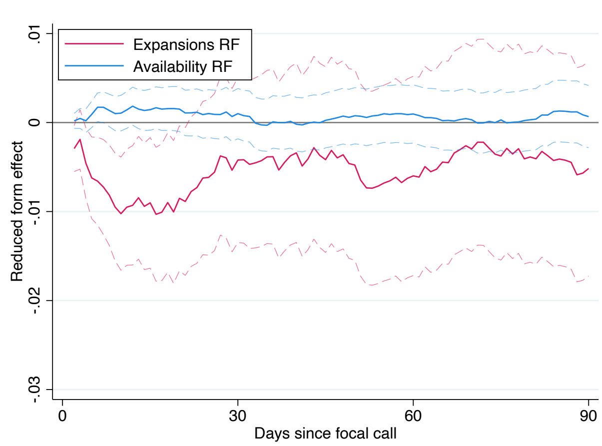
(a) Effect on future 911 call within t days



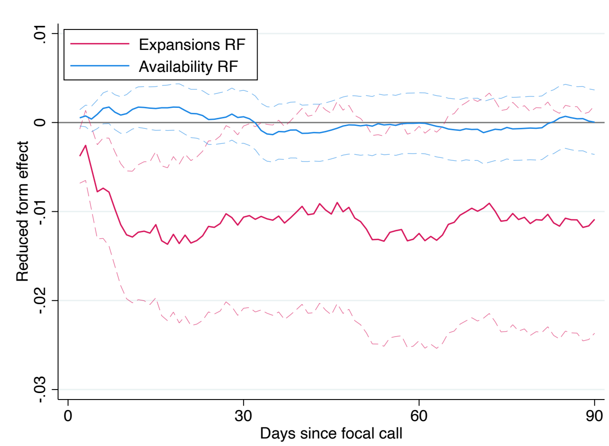
(b) Effect on non-emergency call within t days



(c) Effect on future violent call within t days



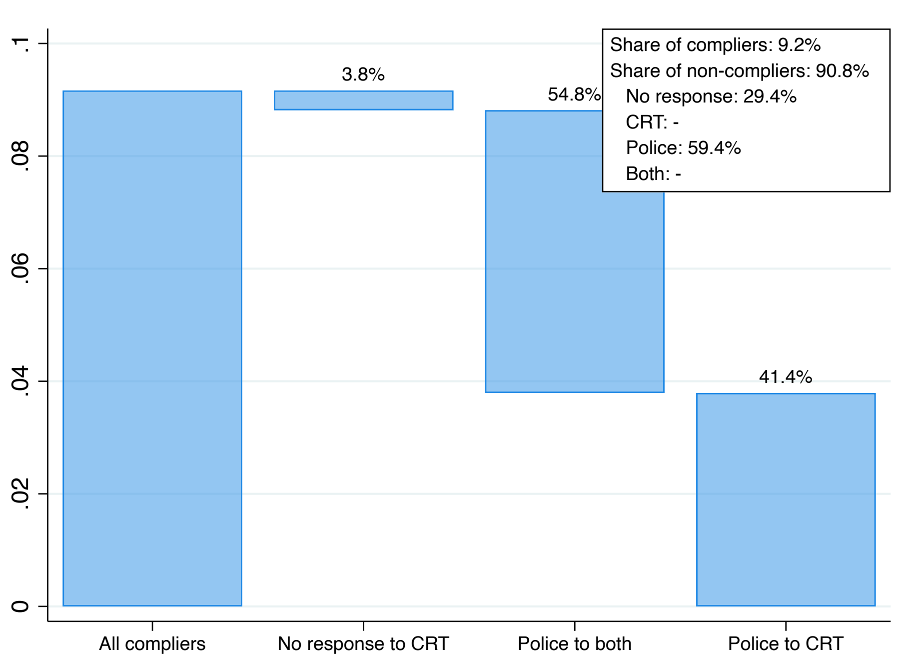
(d) Effect on future serious call within t days



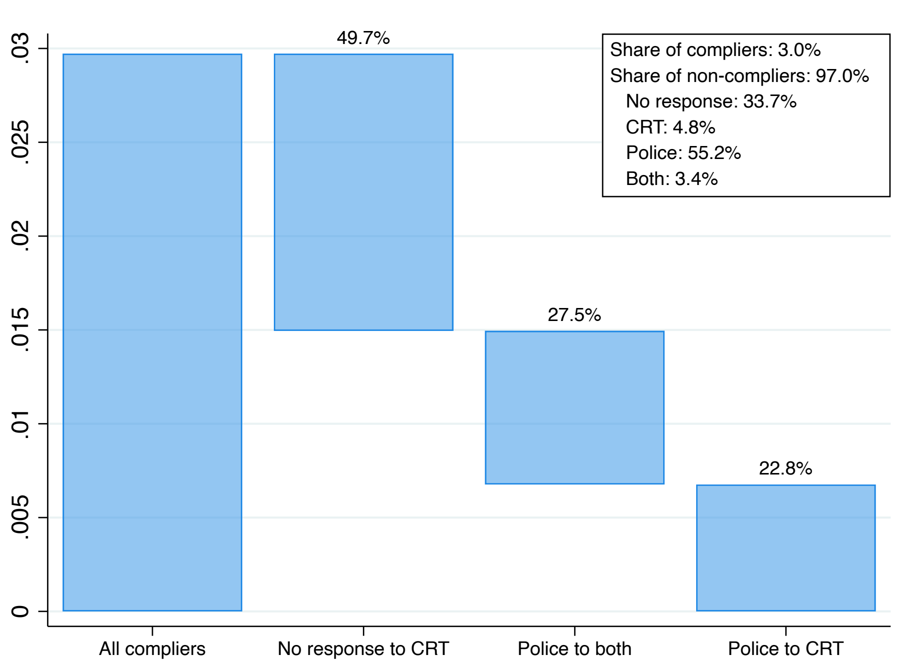
Notes: This figure shows the effect of both the expansion instrument and availability instrument on future 911 calls and non-emergency calls from the same address within a period of time up to 90 days. Each coefficient comes from a regression of an indicator for a type of call occurring within t days on either indicators for each CRT expansion or the availability instrument indicator. Serious calls are defined as calls with violent natures as well as burglaries, hit and runs, and any other nature with an arrest rate higher than the 85th percentile and police response rate higher than 70%. 95% confidence intervals clustered at the individual level are shown in dotted lines.

Figure 6: Shares of response types in the expansions and availability designs

(a) Expansions design



(b) Availability design



Notes: This figure reports the shares of compliers from each response type using the expansions design (top panel) and availability design (bottom panel). For the expansions design, the complier share is estimated using a precision-weighted average of the first-stage relationships between having any CRT response and the five expansion instruments. For the availability design, the complier share is estimated using the first-stage relationship between an indicator for high CRT capacity and having any CRT response. The complier type shares are estimated using the instrumented impact of any CRT response on an indicator for having any response, for having both police and a CRT respond, or for not having a police response, respectively.

Tables

Table 1: Call characteristics by response

	(1)	(2)	(3)	(4)	(5)	(6)
		Response composition				
	Count	No response	CRT only	Police only	Both	Arrests
All calls	268,658	0.331	0.052	0.581	0.035	0.089
Panel A: Top 5 most frequent natures with a CRT response						
Welfare check	16,083	0.214	0.353	0.369	0.065	0.023
Suicidal person	6,501	0.128	0.243	0.368	0.261	0.084
Traffic/driver issue	31,611	0.486	0.023	0.482	0.010	0.042
Disorderly subject	19,953	0.206	0.044	0.683	0.067	0.146
Intoxicated person	1,393	0.202	0.415	0.314	0.069	0.047
Panel B: Top 5 least frequent natures with a CRT response						
Robbery	865	0.091	0.001	0.881	0.027	0.338
Fight	1,680	0.092	0.000	0.893	0.015	0.138
Drunk driver	9,707	0.621	0.000	0.377	0.002	0.036
Warrant	1,445	0.001	0.000	0.971	0.028	0.988
Armed subject	1,403	0.100	0.001	0.870	0.028	0.207
Panel C: Top 5 most frequent natures with a police response						
Traffic/driver issue	31,611	0.486	0.023	0.482	0.010	0.042
Disorderly subject	19,953	0.206	0.044	0.683	0.067	0.146
Suspicious person	17,002	0.241	0.013	0.722	0.024	0.047
Mischief/trespassing	13,153	0.321	0.007	0.650	0.022	0.118
Burglary/theft	11,225	0.366	0.001	0.624	0.008	0.146
Panel D: Top 5 most frequent natures with both a police and CRT response						
Dispute	42,838	0.101	0.001	0.863	0.035	0.096
Disorderly subject	19,953	0.206	0.044	0.683	0.067	0.146
Welfare check	16,083	0.214	0.353	0.369	0.065	0.023
Suspicious person	17,002	0.241	0.013	0.722	0.024	0.047
Overdose	3,982	0.418	0.030	0.474	0.078	0.028
Panel E: Change in call nature						
Nature changed	36,694	0.260	0.063	0.621	0.057	0.213

Notes: This table shows average call characteristics by response type for 911 calls in Eugene and Springfield. Data runs from January 1, 2014 to December 14, 2021. Column (1) displays the count of calls. Columns (2)-(5) display the proportion of calls receiving no response, a CRT only response, a police only response, or a joint CRT and police response. Column (6) displays the proportion of calls resulting in an arrest. Panels A and B list call information for the five most frequent call natures receiving a CRT response and the five call natures least likely to receive a CRT response, respectively. Panels C and D list call information for the five most frequent call natures receiving a police response or a joint CRT and police response, respectively. Panel E shows the proportion of calls with each response type given they had a nature change.

Table 2: Call outcomes by response

	(1) Count	(2) CRT only	(3) Police only	(4) Both
Panel A: Call response				
Call time to dispatch time (min.)	268,658	50.47	17.96	11.58
Dispatch time to arrival time (min.)	268,658	13.70	6.31	6.55
Police time on call (min.)	268,658	0.00	52.32	72.64
CRT time on call (min.)	268,658	24.34	0.00	41.26
Police arrives first	268,658	0.000	1.000	0.882
CRT arrives first	268,658	1.000	0.000	0.116
Panel B: Call escalation outcomes				
Arrest	268,658	0.000	0.145	0.130
Detox	268,658	0.000	0.002	0.006
Non-criminal hold	268,658	0.000	0.010	0.040
Warrant	268,658	0.000	0.027	0.013
Invol. hold	268,658	0.000	0.012	0.046
Use of force	268,658	0.000	0.008	0.011
Other arrest	268,658	0.000	0.106	0.072
Panel C: Report outcomes				
Any report	268,658	0.000	0.279	0.282
Mental health	268,658	0.000	0.012	0.062
Overdose	268,658	0.000	0.005	0.011
Any criminal activity	268,658	0.000	0.211	0.147
Violent	268,658	0.000	0.048	0.051
Drug	268,658	0.000	0.005	0.003
Property	268,658	0.000	0.052	0.013

Notes: This table shows average call outcomes by response type for 911 calls in Eugene and Springfield. Data runs from January 1, 2014 to December 14, 2021. Column (1) displays the count of calls. Columns (2)-(4) display the proportion of calls receiving a CRT only response, a police only response, or a joint CRT and police response. Panel A displays call response outcomes. Panel B displays arrest outcomes. Panel C displays report outcomes.

Table 3: Effects of CRTs on call response and arrest using the expansions design

	Call responses			Call outcomes			
	CRT	Police	Any response	Arrest	Invol. hold	Detox	Warrant
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Reduced form effect of expansion						
Post expansion	0.075***	-0.031***	0.003	-0.018***	-0.011***	-0.004***	0.002
	(0.003)	(0.006)	(0.006)	(0.005)	(0.002)	(0.001)	(0.002)
Untreated mean	0.005	0.722	0.724	0.119	0.018	0.005	0.018
	Panel B: IV estimates of effect of CRT response						
CRT		-0.414***	0.038	-0.243***	-0.151***	-0.052***	0.022
		(0.084)	(0.082)	(0.062)	(0.023)	(0.013)	(0.026)
Control complier mean		0.962	0.962	0.321	0.173	0.055	-0.010
First-stage F-stat		172.60	172.60	172.60	172.60	172.60	172.60
J test of overidentification							
χ^2 statistic		12.305	17.035	1.772	17.360	1.374	6.761
p-value		0.015	0.002	0.778	0.002	0.849	0.149
Hausman test							
χ^2 statistic		3.374	15.677	11.524	58.397	19.394	5.013
p-value		0.066	0.000	0.001	0.000	0.000	0.025
	Panel C: OLS estimates of effect of CRT response						
CRT		-0.227***	0.372***	-0.037***	0.010***	0.001***	-0.013***
		(0.004)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)
Untreated mean		0.639	0.639	0.092	0.006	0.002	0.018
Observations		291,599	291,599	291,599	291,599	291,599	291,599

Notes: This table presents estimates of the impact of a CRT response on call responses and call outcomes using the expansions design. Panel A reports reduced form estimates of the effect of CRT expansions on these call outcomes and responses. Panel B shows 2SLS estimates of the effect of a CRT response on call outcomes and responses using indicators for each expansion as instruments. Panel C shows the association between a CRT response and the same outcomes in the last two panels using the same set of controls. We use a stacked design that combines all policy changes as instruments for a CRT response. Heteroskedasticity robust standard errors, clustered by call, are in parentheses. Significance stars indicate * p < 0.10, ** p < 0.05, *** p < 0.01

Table 4: Effects of CRTs on call response and arrest using the availability design

	Call responses			Call outcomes			
	CRT (1)	Police (2)	Any response (3)	Arrest (4)	Invol. hold (5)	Detox (6)	Warrant (7)
Panel A: Reduced form effect of high CRT capacity							
High CRT capacity (=1)	0.030***	-0.007***	0.015***	-0.001	-0.000	0.000**	-0.001
	(0.001)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(0.001)
Untreated mean	0.083	0.607	0.654	0.086	0.006	0.001	0.016
Panel B: IV estimates of effect of CRT response							
CRT		-0.228***	0.497***	-0.045	-0.004	0.012**	-0.020
		(0.071)	(0.067)	(0.041)	(0.011)	(0.005)	(0.018)
Control complier mean		0.503	0.503	0.080	0.025	-0.006	0.017
First-stage F-stat		522.67	522.67	522.67	522.67	522.67	522.67
Hausman test							
χ^2 statistic		0.012	3.311	0.086	2.270	4.638	0.204
p-value		0.912	0.069	0.769	0.132	0.031	0.652
Equality with DinD LATE (p)		0.091	0.000	0.008	0.000	0.000	0.200
Panel C: OLS estimates of effect of CRT response							
CRT		-0.220***	0.381***	-0.033***	0.012***	0.001***	-0.012***
		(0.004)	(0.002)	(0.002)	(0.001)	(0.000)	(0.001)
Untreated mean		0.624	0.624	0.088	0.005	0.001	0.017
Observations		236,534	236,534	236,534	236,534	236,534	236,534

Notes: This table presents estimates of the impact of a CRT response on call responses and call outcomes using the availability design. Panel A reports reduced form estimates using an indicator for high CRT capacity when focal call comes in on call outcomes and responses. Panel B shows 2SLS estimates of the effect of a CRT response on call outcomes and responses using the CRT capacity indicator as an instrument for having a CRT response. Panel C shows the association between a CRT response and the same outcomes in the last two panels using the same set of controls. Heteroskedasticity robust standard errors are in parentheses. Significance stars indicate * p < 0.10, ** p < 0.05, *** p < 0.01

Table 5: Potential response types for a binary instrument that shifts CRT availability

Panel A: Expansions design

		After expansion			
		No response (n)	Only CRT (c)	Only police (p)	Both (b)
Before expansion	No response (n)	π_{nn}	π_{nc}	-	-
	Only police (p)	-	π_{pc}	π_{pp}	π_{pb}

Panel B: Availability design

		More availability			
		No response (n)	Only CRT (c)	Only police (p)	Both (b)
Less availability	No response (n)	π_{nn}	π_{nc}	-	-
	Only CRT (c)	-	π_{cc}	-	-
	Only police (p)	-	π_{pc}	π_{pp}	π_{pb}
	Both (b)	-	-	-	π_{bb}

Notes: This table presents the potential response types allowed under our choice model for each research design, where $\pi_{xy} \equiv \Pr[D(0){=}d, D(1){=}d']$. Blue shading denotes response types ruled out by Assumption 2 (monotonicity), while orange shading denotes response types ruled out by Assumption 1 (no cross effects except substitution). Panel A presents the response types for the expansions design, and Panel B for the availability design. In the expansions design, the set of feasible response types is more limited because CRTs were not available prior to the reform; as a result, response types such as $c \rightarrow p$ are ruled out by construction.

Table 6: Cross-design comparison of complier mean arrest rates

	Design	
	Availability	Diff-in-diff
	(1)	(2)
Panel A: Comparing $E[Y(b)]$ among $p{\to}b$ compliers		
$E[Y(b) \mid p{\to}b]$	0.128***	0.142***
	(0.034)	(0.016)
p-value of equality		(0.688)
J test of overidentification		
χ^2 statistic		0.530
p-value		0.971
Panel B: Comparing $E[Y(p)]$ among $p{\to}b$ and $p{\to}c$ compliers		
$\frac{s_{pb}}{s_{pc}+s_{pb}}E[Y(p) \mid p{\to}b] + \frac{s_{pc}}{s_{pc}+s_{pb}}E[Y(p) \mid p{\to}c]$	0.159**	0.321***
	(0.074)	(0.064)
p-value of equality		(0.098)
$s_{pb}/(s_{pc}+s_{pb})$	0.548	0.570
$s_{pc}/(s_{pc}+s_{pb})$	0.452	0.430
Panel C: Treatment effect estimates		
Rescaled CRT effect ($\frac{s_{pc}}{s_{pc}+s_{pb}}\Delta_{pc} + \frac{s_{pb}}{s_{pc}+s_{pb}}\Delta_{pb}$)	-0.246***	-0.089
	(0.064)	(0.077)

Notes: This table presents estimates of complier mean arrest rates in each of the research designs. Panel A shows an estimate of $E[Y(b)|p \to b]$ using the first moment in Equation 6. Panel B shows the weights and the estimated moment in the second equation in Equation 6. Significance stars indicate * p < 0.10, ** p < 0.05, *** p < 0.01

Table 7: Complier group-specific treatment effects and outcome means for arrest

	(1)
Panel A: Complier means	
$E[Y(b) \mid p\to b]$	0.141***
	(0.013)
$E[Y(p) \mid p\to b]$	0.498***
	(0.174)
$E[Y(p) \mid p\to c]$	0.120
	(0.184)
Panel B: subLATEs	
Police $\to$ both	-0.357**
	(0.174)
Police $\to$ CRT	-0.120
	(0.184)
Panel C: Goodness-of-fit tests	
J test of overidentification	
χ^2 statistic	0.952
p-value	0.996
J test of overidentification for $E[Y(p)]$ moment	
χ^2 statistic	0.422
p-value	0.936

Notes: This table reports estimates of the complier means and associated treatment effects. Our approach is to estimate the complier means $E[Y(d) \mid g]$ that best fit the moments

$$\frac{E[Y1[D=b \mid Z=1]] - E[Y1[D=b \mid Z=0]]}{E[1[D=b \mid Z=1]] - E[1[D=b \mid Z=0]]} = E[Y(b) \mid p\to b]$$

$$\frac{E[Y1[D=p \mid Z=1]] - E[Y1[D=p \mid Z=0]]}{E[1[D=p \mid Z=1]] - E[1[D=p \mid Z=0]]} = \frac{s_{pb}}{s_{pb}+s_{pc}} E[Y(p) \mid p\to b] + \frac{s_{pb}}{s_{pb}+s_{pc}} E[Y(p) \mid p\to c].$$

Since there are five expansions, we have an overidentified set of equations with ten moments and three unknowns. We variance-weight the moments using a diagonal weighting matrix and solve via minimum distance. We also report J tests of overidentification for all moments together, as well as for the second moment (the J test for the $E[Y(b) \mid p\to b]$ moment can be found in Table 6). Significance stars indicate * p < 0.10, ** p < 0.05, *** p < 0.01

Table 8: Marginal Value of Public Funds (MVPF)

	Average Response at Scale	Marginal Expansion
Costs per Response		
Average cost per response	\$253.88	\$253.88
Total cost	\$759,853.61	\$759,853.61
Number of CAHOOTS responses	2,993	2,993
Fiscal Externalities		
Substitution (ΔPolice Response $\times$ 931)	\$-385.43	\$-212.27
Fewer arrests (ΔArrests $\times$ 707)	\$-171.80	\$-31.82
Fiscal externality	-\$557.24	-\$244.08
Cost + fiscal externalities	-\$303.36	\$9.79
Additional crime reductions		
Fiscal externality from crime reductions	-\$391.55	-\$45.50
Cost + all fiscal externalities	-\$694.84	-\$35.63
Per-adult Δ in victim costs	-\$22.66	-\$3.04
MVPF scenarios		
Full cost + base fiscal externalities		
Per-adult WTP is \$1	∞	5.09
Per-adult WTP is \$69 (Ba et al., 2025)	∞	351.31
Assume marginal response \$10 cheaper than average response		
Per-adult WTP is positive	∞	∞
Incorporating crime reductions		
Per-adult WTP is -Δ victim costs	∞	∞

Notes: This table shows estimates of the MVPF. We estimate the cost per response by dividing the total annual cost of CAHOOTS in Eugene by the number of responses. We estimate fiscal externalities from substitution between CRTs and police and from fewer arrests by scaling our estimated effects on these outcomes by the cost of a police response or of an arrest. EPD reported the cost of a police response was over \$800 in 2013 which is equivalent to \$931 in 2021. We assume arrests are \$707 based on the cost of a mental health hold reported for the Northwest in Karaca and Moore (2020), Table 2. The additional crime reductions combine our estimates in Table A13 with the enforcement, court, and victim cost estimates in Deshpande and Mueller-Smith (2022) for assault, theft, and drug possession. All dollars converted to 2021 dollars using CPI-U.

A1 The effect of relative arrival time in responses with both police and a CRT

One important question that remains is the extent to which the first-arriving unit sets the stage for the rest of the emergency response. In principle, a CRT arriving first might change the tenor of the initial interaction and further reduce the likelihood of an arrest. This is particularly important since the vast majority of b calls in our context involve the police arriving first, and we expect that if the CRT program was expanded it would be feasible for CRTs to arrive first in a larger share of calls.

To explore the importance of relative arrival time, one natural experiment would be to focus on calls where both the police and a CRT were simultaneously dispatched, but idiosyncratic differences in their current location and in local traffic conditions created quasi-exogenous variation in which unit arrived first.

In our context, this analysis is infeasible since we do not observe the location of each unit over time. To try to zoom in on the same type of calls, we instead focus on calls where both types of units were dispatched nearly simultaneously and well before the first arrival. In particular, we start with the set of calls where both CRT and police units were dispatched within two minutes of each other, and the last dispatch time occurred at least five minutes before the first unit arrival. We then regress outcomes on a dummy for the police arriving at least two minutes before a CRT, a dummy for a CRT arriving at least two minutes before the police, a control for the relative time of dispatch, and fixed effects at the city by year and city by hour of day level. Thus, the coefficients reveal the relative effect of an early arrival by one of the types of units relative to an almost-simultaneous arrival.

Table A14 displays the results, and reveals several interesting facts. First, there are only 704 such calls, less than 10% of the 9,494 calls that have both a CRT and police response. Consistent with the evidence we discussed earlier (and with CAHOOTS not being a co-response model), this tells us that simultaneous dispatch of both types of units is relatively rare.

Second, the relative arrival time is not correlated with the arrest rates as predicted from the tract characteristics (column 1), suggesting that the arrival time is quasi-exogenous conditional on these controls. However, it does not appear that the arrival order has a substantive effect on the call outcome, with both of the early arrival coefficients small and statistically insignificant across specifications. While we view this evidence as merely suggestive—since it implicitly conditions on factors that happen after the initial dispatch that lead to both types of units ever arriving—it is consistent with the relative arrival time not having a large impact on call outcomes, and implies that little is lost by not distinguishing between different types of b calls.

A2 Econometric Appendix

Given the four treatments, there are eight moments that reflect the effect of the instrument on the outcome multiplied by an indicator for receiving or not receiving a treatment d, rescaled by the relevant first stage. These moments are as follows:

$$\begin{bmatrix} \frac{E[Y1[D{=}b|Z{=}1]]-E[Y1[D{=}b|Z{=}0]]}{E[1[D{=}b|Z{=}1]]-E[1[D{=}b|Z{=}0]]} \\ \frac{E[Y1[D{\neq}b|Z{=}1]]-E[Y1[D{\neq}b|Z{=}0]]}{E[1[D{\neq}b|Z{=}1]]-E[1[D{\neq}b|Z{=}0]]} \\ \frac{E[Y1[D{=}c|Z{=}1]]-E[Y1[D{=}c|Z{=}0]]}{E[1[D{=}c|Z{=}1]]-E[1[D{=}c|Z{=}0]]} \\ \frac{E[Y1[D{\neq}c|Z{=}1]]-E[Y1[D{\neq}c|Z{=}0]]}{E[1[D{\neq}c|Z{=}1]]-E[1[D{\neq}c|Z{=}0]]} \\ \frac{E[Y1[D{=}p|Z{=}1]]-E[Y1[D{=}p|Z{=}0]]}{E[1[D{=}p|Z{=}1]]-E[1[D{=}p|Z{=}0]]} \\ \frac{E[Y1[D{\neq}p|Z{=}1]]-E[Y1[D{\neq}p|Z{=}0]]}{E[1[D{\neq}p|Z{=}1]]-E[1[D{\neq}p|Z{=}0]]} \\ \frac{E[Y1[D{=}n|Z{=}1]]-E[Y1[D{=}n|Z{=}0]]}{E[1[D{=}n|Z{=}1]]-E[1[D{=}n|Z{=}0]]} \\ \frac{E[Y1[D{\neq}n|Z{=}1]]-E[Y1[D{\neq}n|Z{=}0]]}{E[1[D{\neq}n|Z{=}1]]-E[1[D{\neq}n|Z{=}0]]} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{-\pi_{pc}}{\pi_{pb}} & \frac{\pi_{pc}}{\pi_{pb}} & \frac{-\pi_{nc}}{\pi_{pb}} & \frac{\pi_{nc}}{\pi_{pb}} \\ 0 & 0 & \frac{\pi_{pc}}{\pi_{pc}+\pi_{nc}} & 0 & \frac{\pi_{nc}}{\pi_{pc}+\pi_{nc}} & 0 \\ \frac{-\pi_{pb}}{\pi_{pc}+\pi_{nc}} & \frac{\pi_{pb}}{\pi_{pc}+\pi_{nc}} & 0 & \frac{\pi_{pc}}{\pi_{pc}+\pi_{nc}} & 0 & \frac{\pi_{nc}}{\pi_{pc}+\pi_{nc}} \\ 0 & \frac{\pi_{pb}}{\pi_{pb}+\pi_{pc}} & 0 & \frac{\pi_{pc}}{\pi_{pb}+\pi_{pc}} & 0 & 0 \\ \frac{\pi_{pb}}{\pi_{pb}+\pi_{pc}} & 0 & \frac{\pi_{pc}}{\pi_{pb}+\pi_{pc}} & 0 & \frac{\pi_{nc}}{\pi_{pb}+\pi_{pc}} & \frac{-\pi_{nc}}{\pi_{pb}+\pi_{pc}} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \frac{\pi_{pb}}{\pi_{nc}} & \frac{-\pi_{pb}}{\pi_{nc}} & \frac{\pi_{pc}}{\pi_{nc}} & \frac{-\pi_{pc}}{\pi_{nc}} & \frac{\pi_{nc}}{\pi_{nc}} & 0 \end{bmatrix} \begin{bmatrix} E[Y(b) \mid p{\to}b] \\ E[Y(p) \mid p{\to}b] \\ E[Y(c) \mid p{\to}c] \\ E[Y(p) \mid p{\to}c] \\ E[Y(c) \mid n{\to}c] \\ E[Y(n) \mid n{\to}c] \end{bmatrix}$$

Note that:

$$\frac{E[Y1[D{=}d] \cdot 1[Z{=}z]]|Z{=}1] - E[Y1[D{=}d] \cdot 1[Z{=}z]|Z{=}0]}{E[1[D{=}d]|Z{=}1] - E[1[D{=}d]|Z{=}0]} = E[Y|D = d, Z = z]$$

and therefore the moments we express above in the form of Wald estimators can be mapped to $E[Y|D = d, Z = z]$.

Given that in our setting, $E[Y(c) \mid p{\to}c] = E[Y(c) \mid n{\to}c] = E[Y(n) \mid n{\to}c] = 0$, this expression can be simplified to

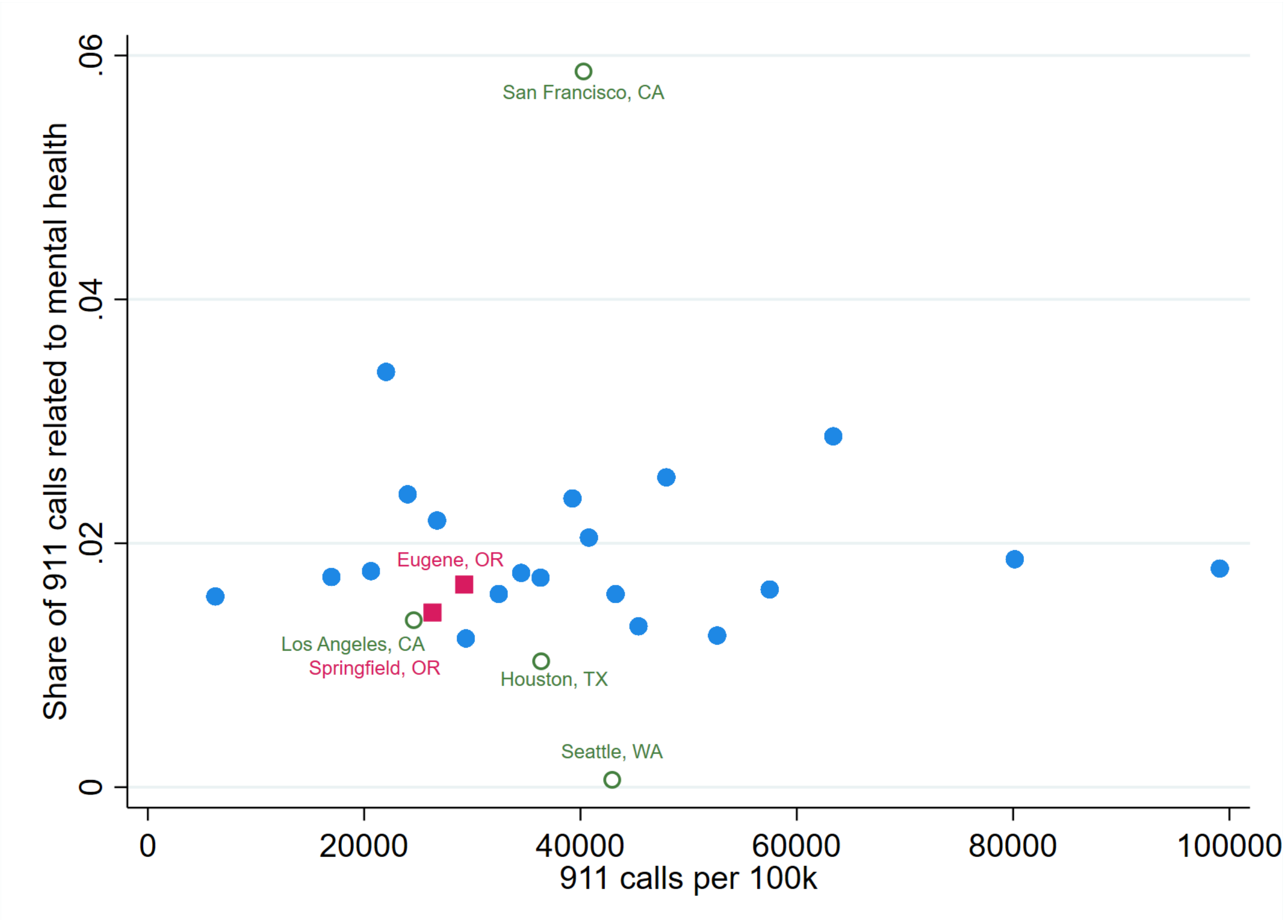
$$\begin{bmatrix} \frac{E[Y1[D{=}b|Z{=}1]]-E[Y1[D{=}b|Z{=}0]]}{E[1[D{=}b|Z{=}1]]-E[1[D{=}b|Z{=}0]]} \\ \frac{E[Y1[D{\neq}b|Z{=}1]]-E[Y1[D{\neq}b|Z{=}0]]}{E[1[D{\neq}b|Z{=}1]]-E[1[D{\neq}b|Z{=}0]]} \\ \frac{E[Y1[D{=}c|Z{=}1]]-E[Y1[D{=}c|Z{=}0]]}{E[1[D{=}c|Z{=}1]]-E[1[D{=}c|Z{=}0]]} \\ \frac{E[Y1[D{\neq}c|Z{=}1]]-E[Y1[D{\neq}c|Z{=}0]]}{E[1[D{\neq}c|Z{=}1]]-E[1[D{\neq}c|Z{=}0]]} \\ \frac{E[Y1[D{=}p|Z{=}1]]-E[Y1[D{=}p|Z{=}0]]}{E[1[D{=}p|Z{=}1]]-E[1[D{=}p|Z{=}0]]} \\ \frac{E[Y1[D{\neq}p|Z{=}1]]-E[Y1[D{\neq}p|Z{=}0]]}{E[1[D{\neq}p|Z{=}1]]-E[1[D{\neq}p|Z{=}0]]} \\ \frac{E[Y1[D{=}n|Z{=}1]]-E[Y1[D{=}n|Z{=}0]]}{E[1[D{=}n|Z{=}1]]-E[1[D{=}n|Z{=}0]]} \\ \frac{E[Y1[D{\neq}n|Z{=}1]]-E[Y1[D{\neq}n|Z{=}0]]}{E[1[D{\neq}n|Z{=}1]]-E[1[D{\neq}n|Z{=}0]]} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{\pi_{pc}}{\pi_{pb}} \\ 0 & 0 & 0 \\ \frac{-\pi_{pb}}{\pi_{pc}+\pi_{nc}} & \frac{\pi_{pb}}{\pi_{pc}+\pi_{nc}} & \frac{\pi_{pc}}{\pi_{pc}+\pi_{nc}} \\ 0 & \frac{\pi_{pb}}{\pi_{pb}+\pi_{pc}} & \frac{\pi_{pc}}{\pi_{pb}+\pi_{pc}} \\ \frac{\pi_{pb}}{\pi_{pb}+\pi_{pc}} & 0 & 0 \\ 0 & 0 & 0 \\ \frac{\pi_{pb}}{\pi_{nc}} & \frac{-\pi_{pb}}{\pi_{nc}} & \frac{-\pi_{pc}}{\pi_{nc}} \end{bmatrix} \begin{bmatrix} E[Y(b) \mid p{\to}b] \\ E[Y(p) \mid p{\to}b] \\ E[Y(p) \mid p{\to}c] \end{bmatrix}$$

Further simplifications are possible. The third and seventh line can be removed because the weight is zero on each potential outcome, the second and fifth line are collinear because the arrest outcome is necessarily zero whenever $D \in \{n, c\}$, the fourth line is a linear combination of the first and fifth line, the sixth line is a rescaled version of the first line, and the eighth line is collinear with the fourth. This leaves only:

$$\begin{bmatrix} \frac{E[Y1[D=b|Z=1]]-E[Y1[D=b|Z=0]]}{E[1[D=b|Z=1]]-E[1[D=b|Z=0]]} \\ \frac{E[Y1[D=p|Z=1]]-E[Y1[D=p|Z=0]]}{E[1[D=p|Z=1]]-E[1[D=p|Z=0]]} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\pi_{pb}}{\pi_{pb}+\pi_{pc}} & \frac{\pi_{pc}}{\pi_{pb}+\pi_{pc}} \end{bmatrix} \begin{bmatrix} E[Y(b) \mid p \to b] \\ E[Y(p) \mid p \to b] \\ E[Y(p) \mid p \to c] \end{bmatrix} \tag{A1}$$

A3 Appendix Figures

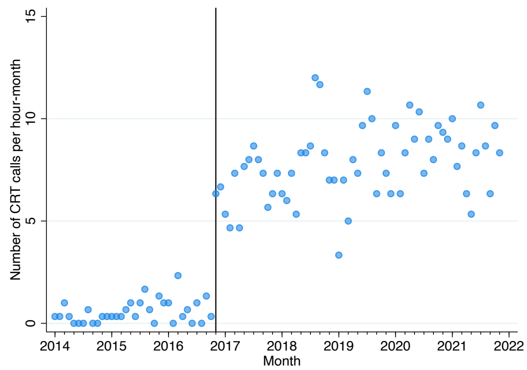
Figure A1: Comparison with other cities, 2018



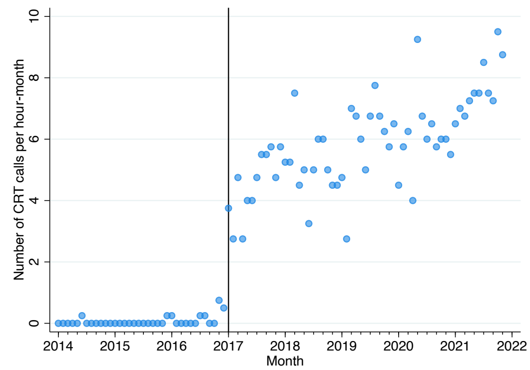
Notes: This figure compares Eugene and Springfield, OR with 134 other cities using a scatter plot of the share of 911 calls that are related to mental health, with particular values highlighted for select cities. Felipe Goncalves provided the data for other cities.

Figure A2: CRT expansions in Eugene and Springfield, 911 calls only

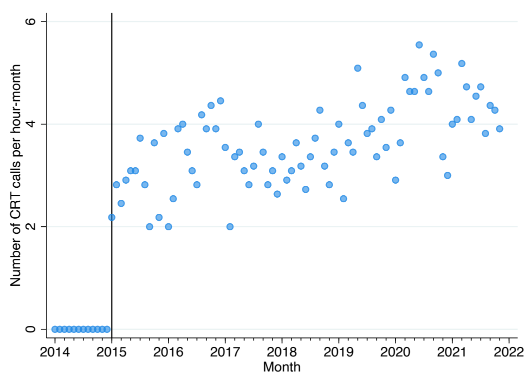
(a) Eugene 7-10 a.m.



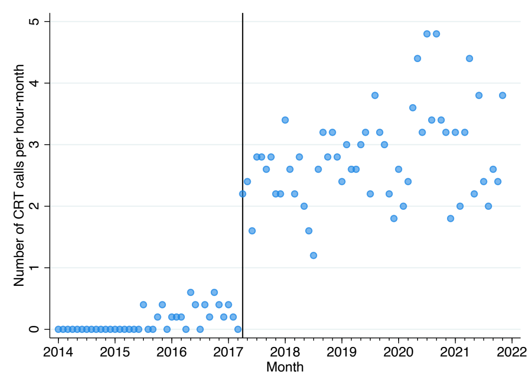
(b) Eugene 3-7 a.m.



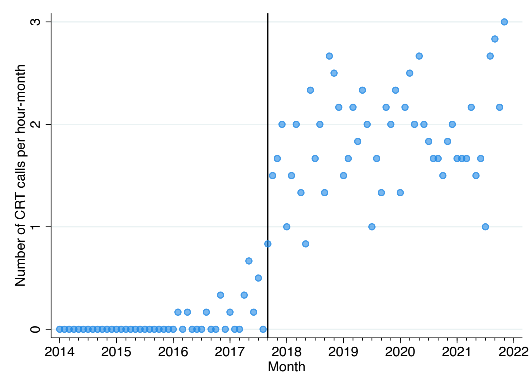
(c) Springfield 12 p.m. to 11 p.m.



(d) Springfield 9-11 a.m. and 11 p.m.-2 a.m.

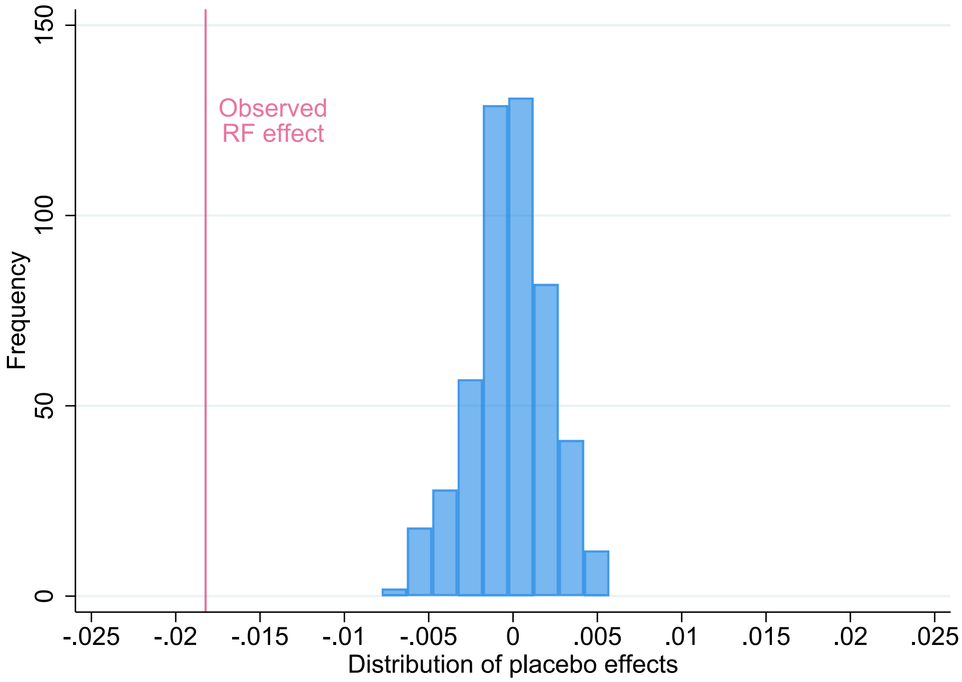


(e) Springfield 3-10 a.m.



Notes: This figure shows the changes in the number of 911 calls only to which a CRT is dispatched as services are expanded (Eugene) or introduced (Springfield). Panels A and B report expansions in Eugene and Panels C, D, and E expansions in Springfield. Each point in the figures reports the number of 911 calls to which a CRT was dispatched, normalized to be measured in terms of calls per hour-month.

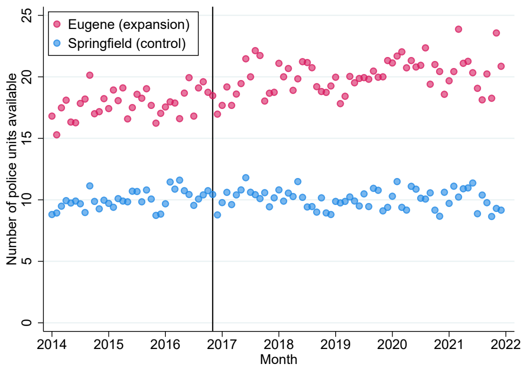
Figure A3: Distribution of placebo expansion times



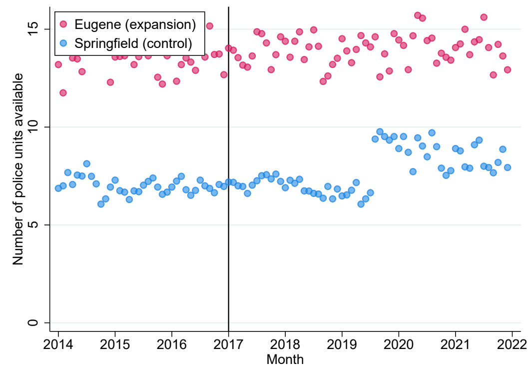
Notes: This figure displays the distribution of estimated effects for 500 randomly chosen event times. These effects are estimated using the same specification as in Equation 3, focusing on the reduced-form relationship. The placebo event time effects are small, centered around zero, and do not overlap with the estimated effect from the observed data.

Figure A4: Police units available during design hours

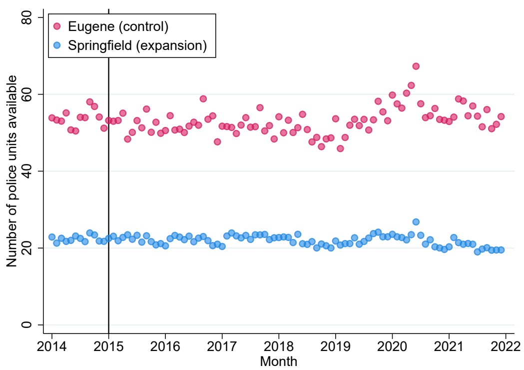
(a) Eugene 7-10 a.m.



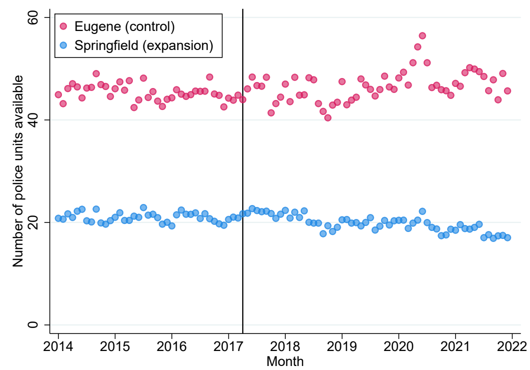
(b) Eugene 3-7 a.m.



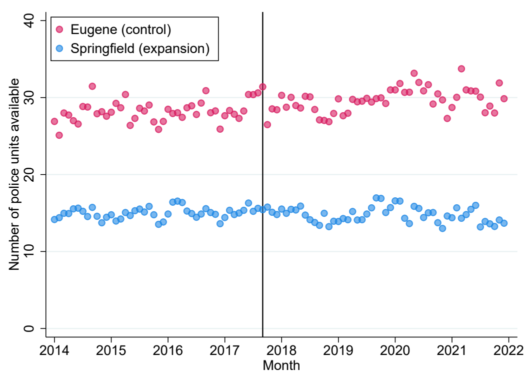
(c) Springfield 12 p.m. to 11 p.m.



(d) Springfield 9-11 a.m. and 11 p.m.-2 a.m.

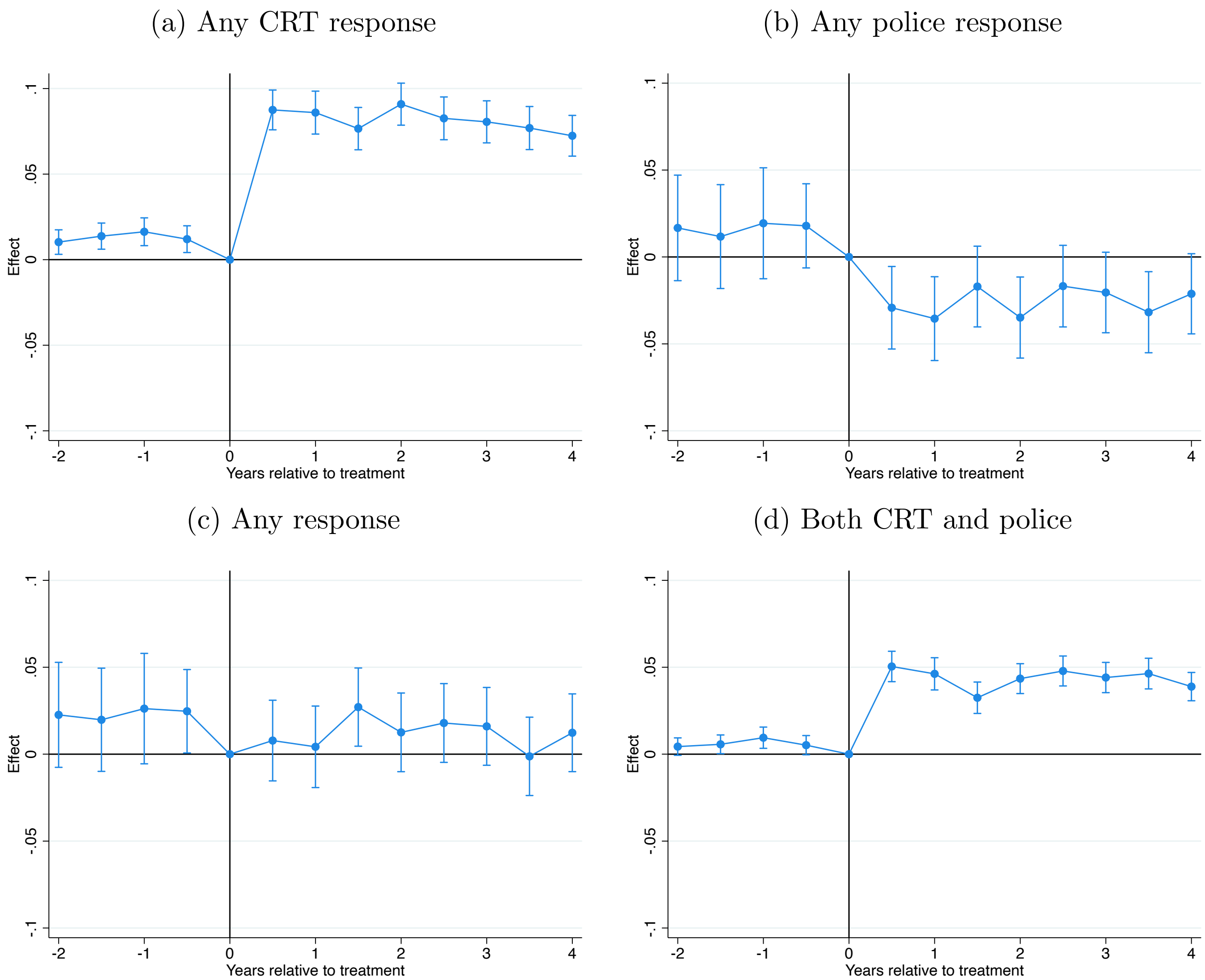


(e) Springfield 3-10 a.m.



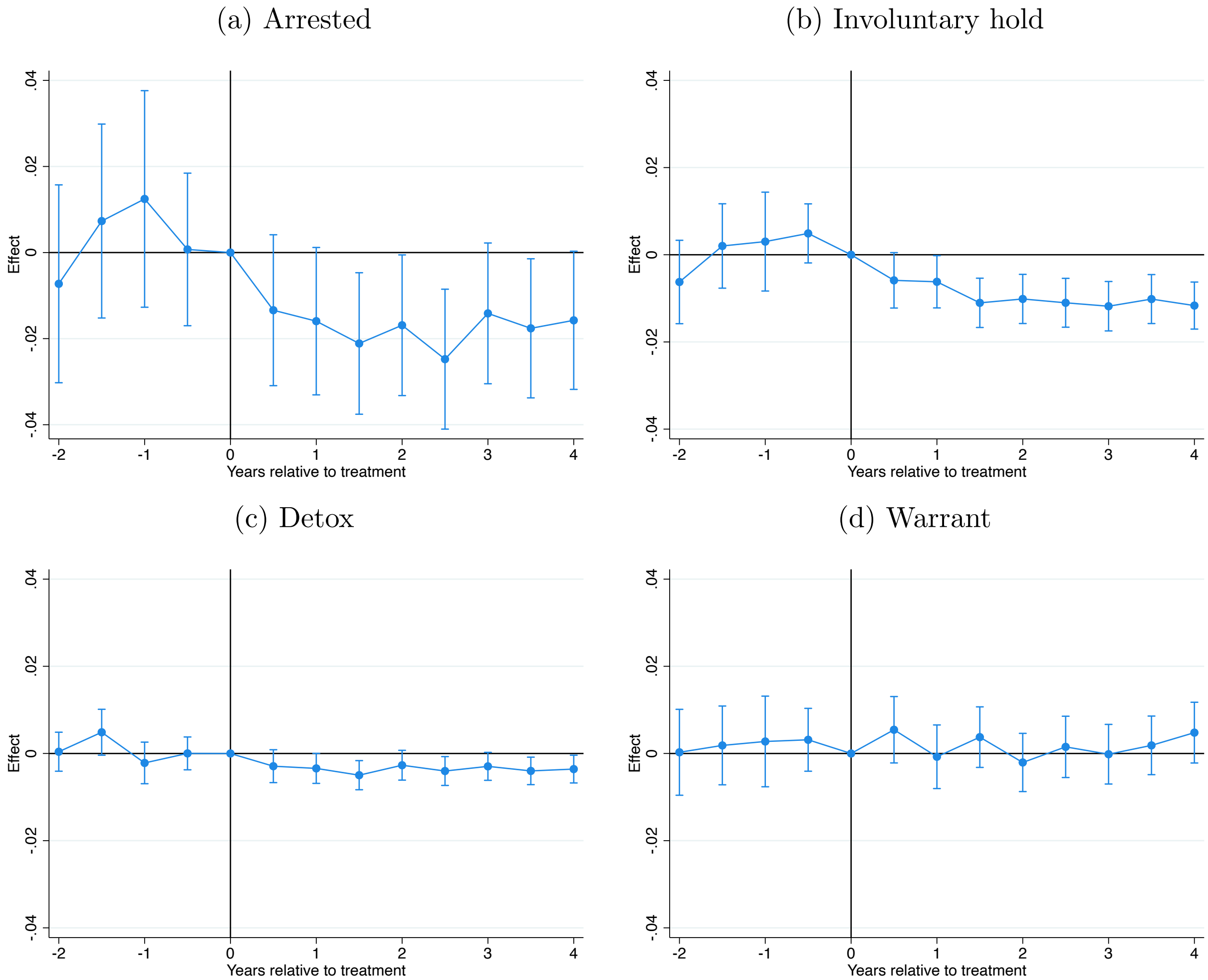
Notes: This figure shows the average number of police units available during design hours in each month and city before and after each expansion. Panels A and B report expansions in Eugene and Panels C, D, and E expansions in Springfield.

Figure A5: Event study estimates of the effect of CRT expansions on type of response



Notes: This figure plots event study coefficients from a regression of call response outcomes on interactions between treatment group indicators (calls from hours and cities affected by each expansion) and six-month time bins relative to the expansion date. The omitted period is months -5 to 0 relative to each expansion. The x-axis shows time in years relative to the expansion, with each point representing a six-month bin. City-design and month-design fixed effects and controls for other policy changes are included. Standard errors clustered at the call level.

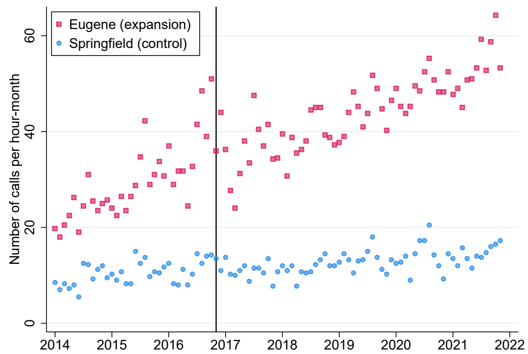
Figure A6: Event study estimates of the effect of CRT expansions on arrests



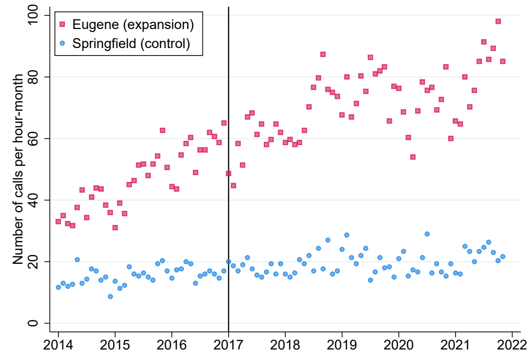
Notes: This figure plots event study coefficients from a regression of arrest outcomes on interactions between treatment group indicators (calls from hours and cities affected by each expansion) and six-month time bins relative to the expansion date. The omitted period is months -5 to 0 relative to each expansion. The x-axis shows time in years relative to the expansion, with each point representing a six-month bin. City-design and month-design fixed effects and controls for other policy changes are included. Standard errors clustered at the call level.

Figure A7: Total call volume by time relative to CRT expansions

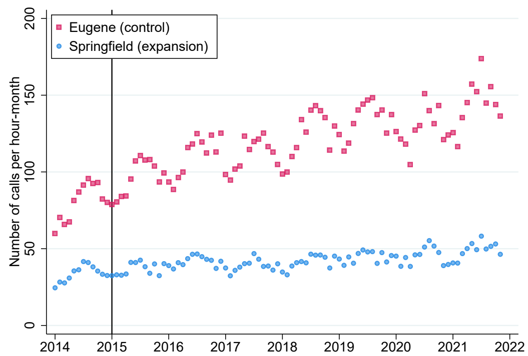
(a) Eugene 7-10 a.m.



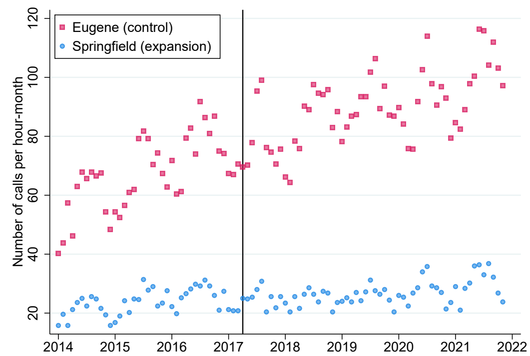
(b) Eugene 3-7 a.m.



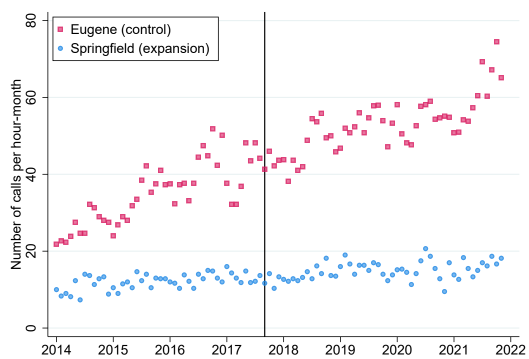
(c) Springfield 12 p.m. to 11 p.m.



(d) Springfield 9-11 a.m. and 11 p.m.-2 a.m.

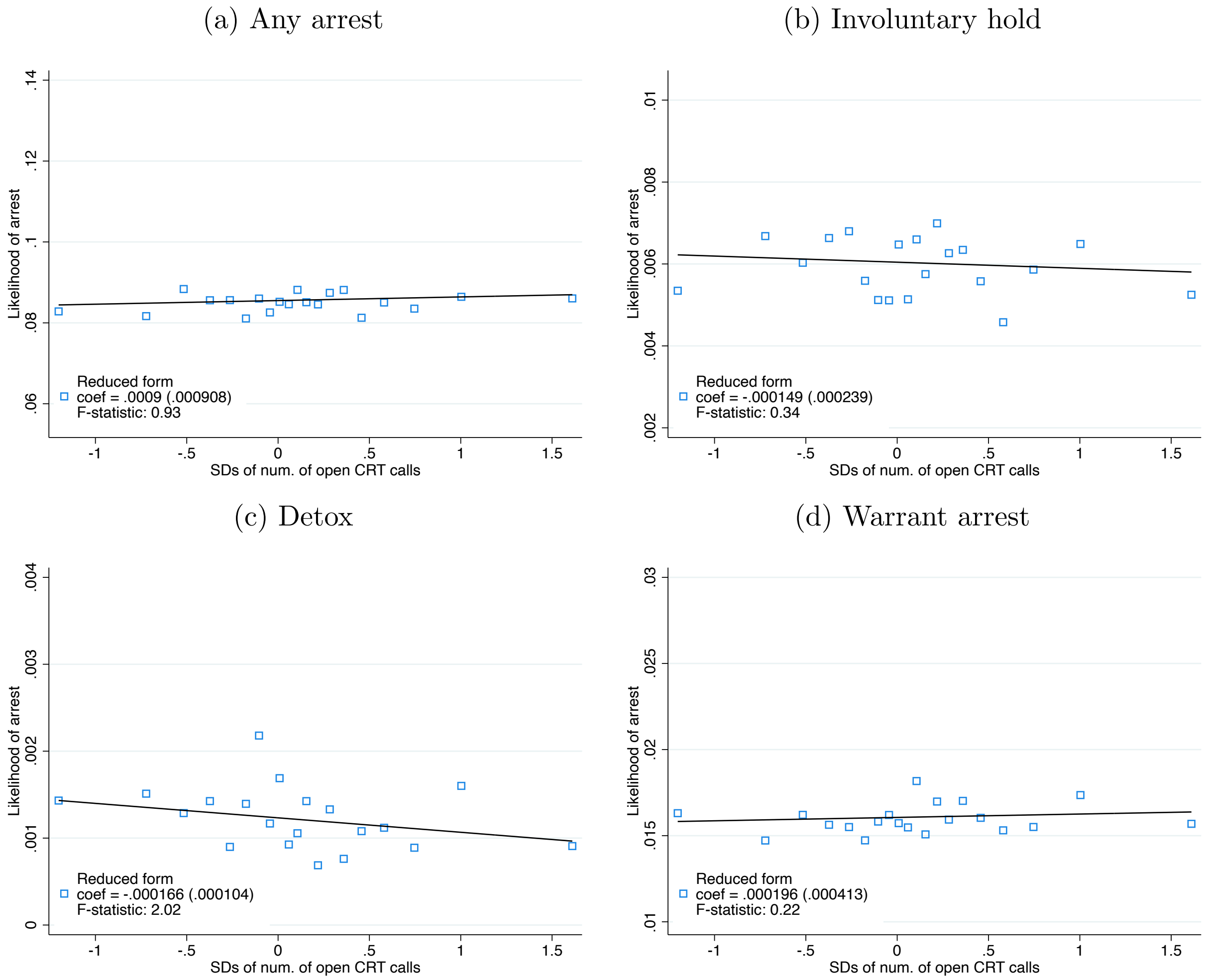


(e) Springfield 3-10 a.m.



Notes: This figure shows the changes in the total number of calls as services are expanded in Eugene and Springfield. Panels A and B report expansions in Eugene and Panels C, D, and E expansions in Springfield. Each point in the figures reports the number of 911 calls, normalized to be measured in terms of calls per hour-month.

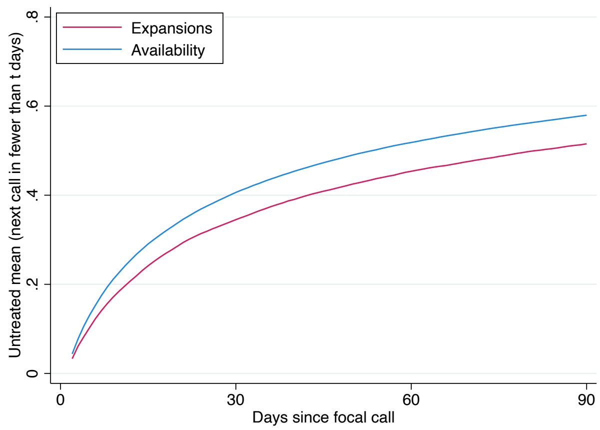
Figure A8: Reduced form of arrest rates on number of open CRT calls



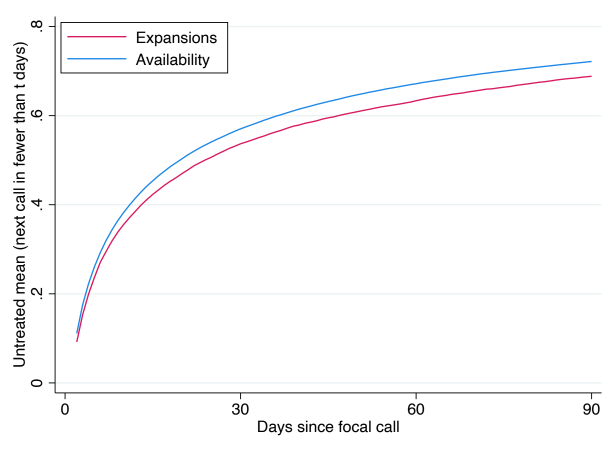
Notes: This figure shows the reduced form relationship between the standardized busyness measure for a CRT and various arrest outcomes including arrests related to involuntary holds, detox and warrants.

Figure A9: Untreated mean of the likelihood of future calls from the same address

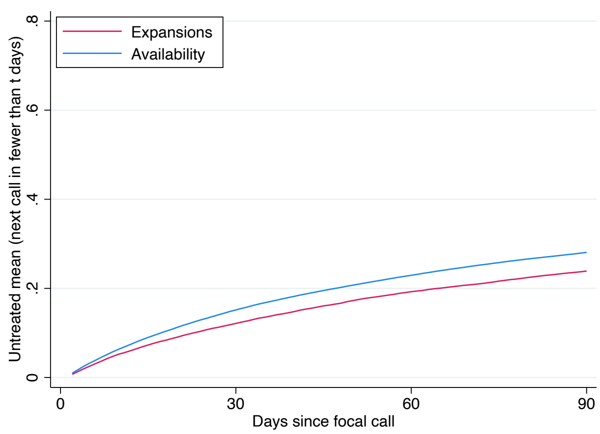
(a) Future 911 call within t days



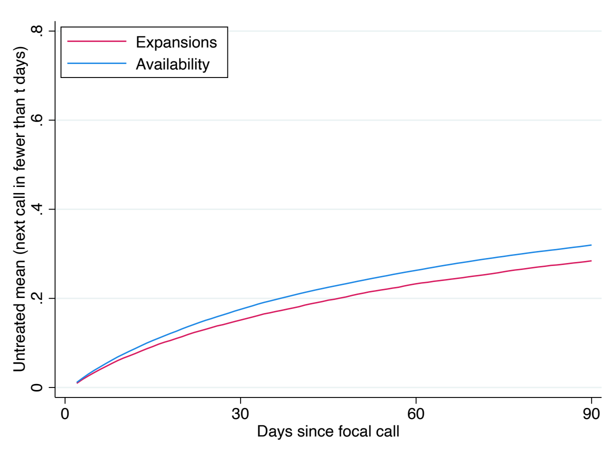
(b) Future non-emergency call within t days



(c) Future violent call within t days



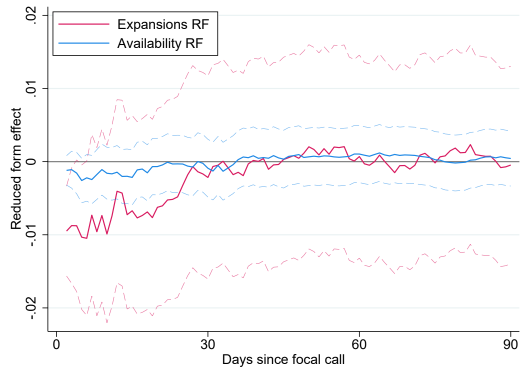
(d) Future serious call within t days



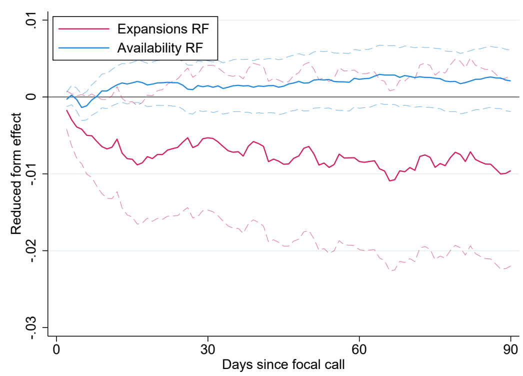
Notes: This figure shows untreated means from a reduced form regression of future calls from the same address within t days on either CRT expansions or the availability indicator. Each panel shows a different type of call.

Figure A10: Robustness for analysis of reduced form effect on future calls from area in fewer than t days

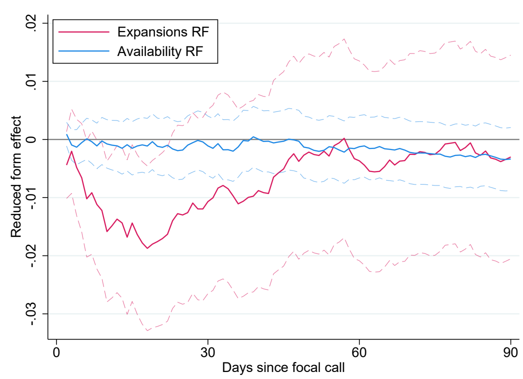
(a) Within 25 meters of focal call



(b) Within 25 meters of focal call, excluding exact location

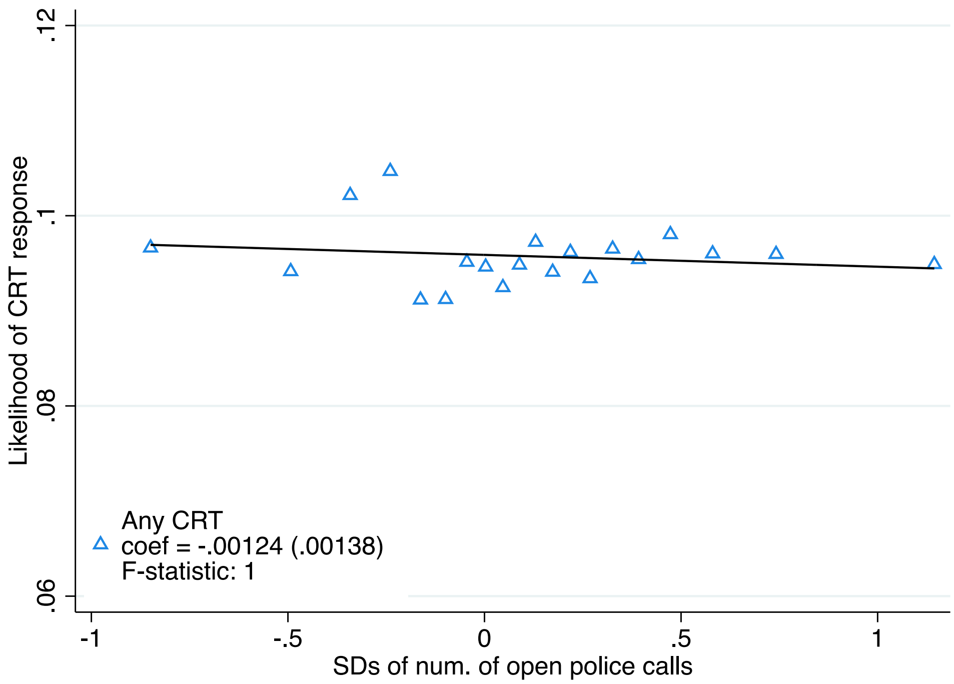


(c) Excluding high-traffic and public places



Notes: Panel A shows the reduced form effect of both instruments on future 911 calls from within 25 meters of the focal call's location within a period of time up to 90 days. Panel B shows analogous results, excluding the exact location of the focal call. Panel C excludes locations with more than 300 calls or with a non-numbered address such as an intersection of a block of a street. Each coefficient comes from a regression of an indicator for a future 911 call within t days on either indicators for each CRT expansion or the availability instrument indicator. 95% confidence intervals clustered at the individual level are shown in dotted lines.

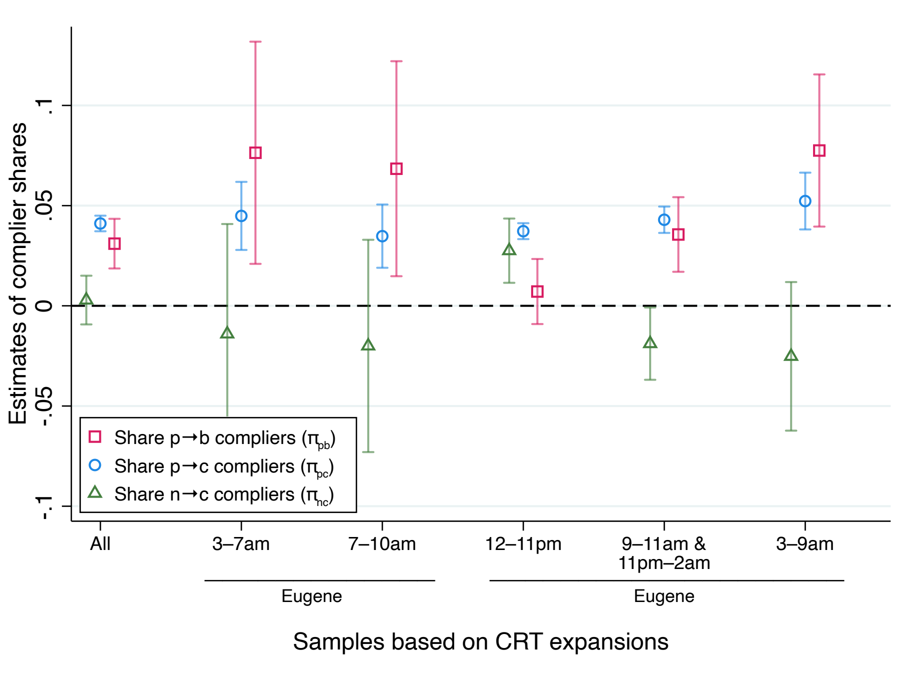
Figure A11: CRT response and police busyness



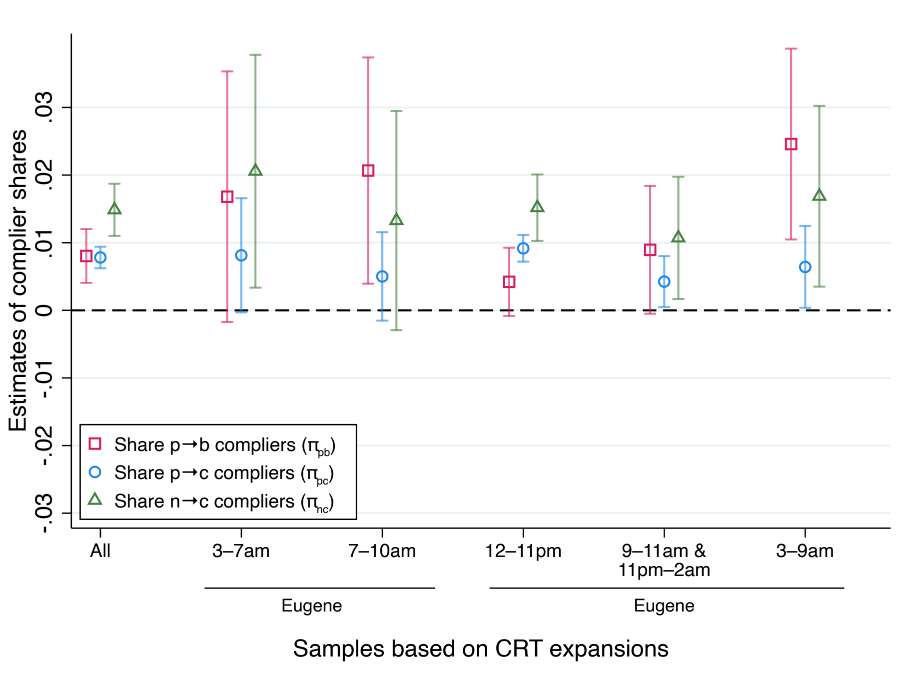
Notes: This figure shows the relationship between a CRT response (y-axis) and police busyness (x-axis). Both variables are residualized with respect to CRT busyness and the fixed effects included in our availability specification (city × month × day-of-week × hour-of-day).

Figure A12: Tests of model assumptions

(a) Expansions design



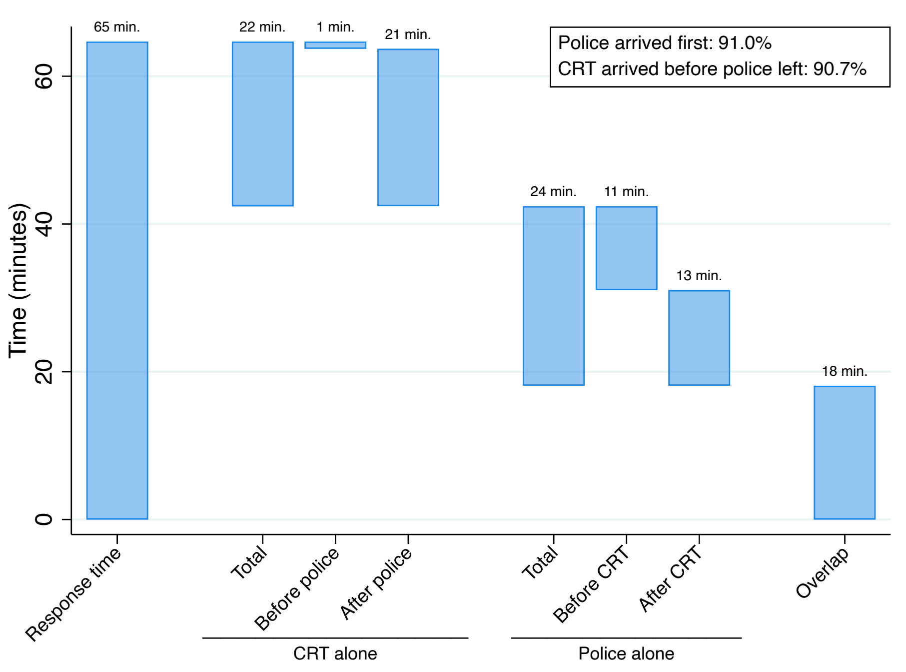
(b) Availability design



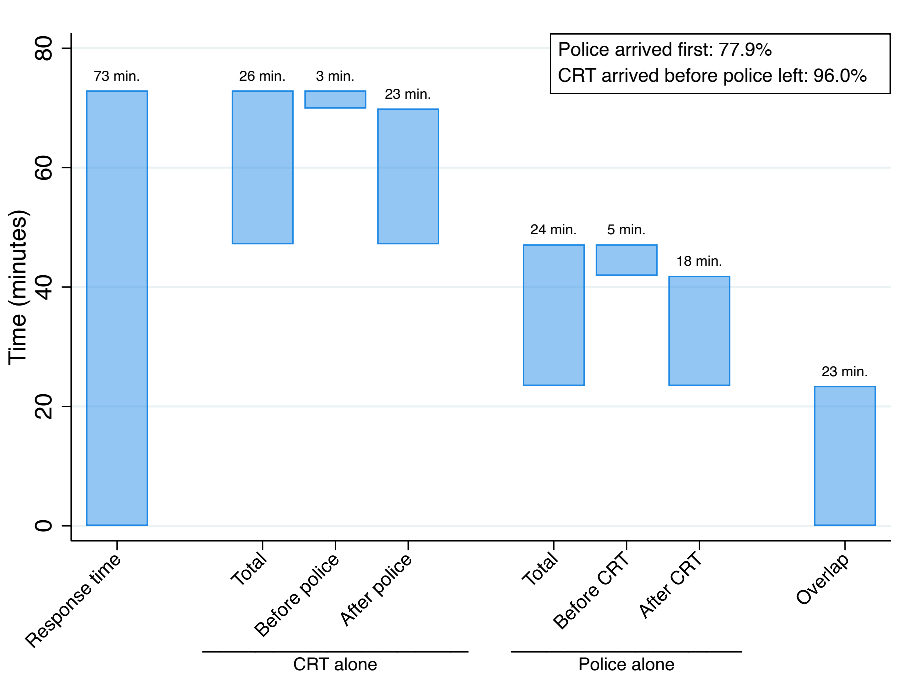
Notes: This figure reports estimates of the shares of the three complier types, π_{pb}, π_{pc}, and π_{nc}, across different sub-samples for each design. The x-axis shows the sub-sample and the y-axis shows the estimated complier share. Sub-samples are defined based on the data used to estimate each specific expansion. Negative estimates of any share indicate potential violations of the model assumptions. Estimation is performed by estimating Equation 5 separately within each sub-sample.

Figure A13: Characterizing complier responses with both police and a CRT

(a) Expansions design



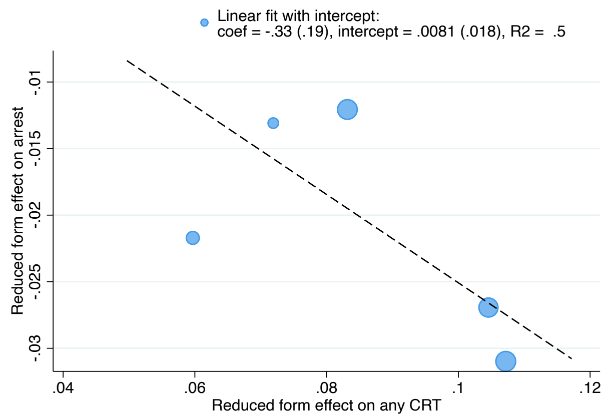
(b) Availability design



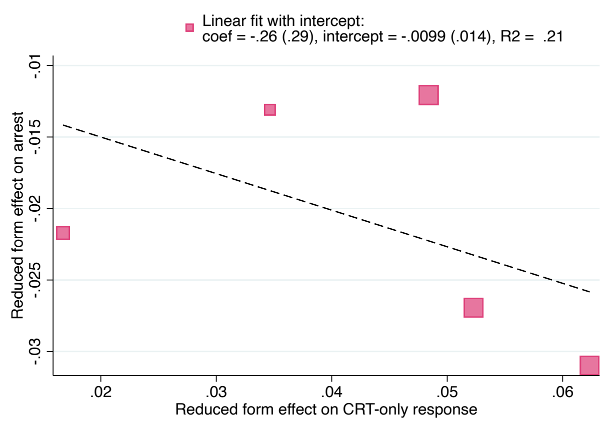
Notes: This figure reports the number of minutes with each type of response among $p \to b$ complier calls. For time measure Y, these times are calculated via a 2SLS regression of $Y 1[D{=}b]$ on $1[D{=}b]$, instrumented with the relevant instrument in each design. Appendix A2 shows that the treated complier means are identified via the following Wald estimator: $\frac{E\left[Y 1[D=b] | Z=1\right] - E\left[Y 1[D=b] | Z=0\right]}{E\left[1[D=b] | Z=1\right] - E\left[1[D=b] | Z=0\right]} = E[Y(b) \mid p \to b]$. Overlap refers to minutes when both a CRT and police were on scene; the minutes alone for each group are broken into time before and after the other unit arrived.

Figure A14: Relationship between effects on arrests and effects on response type

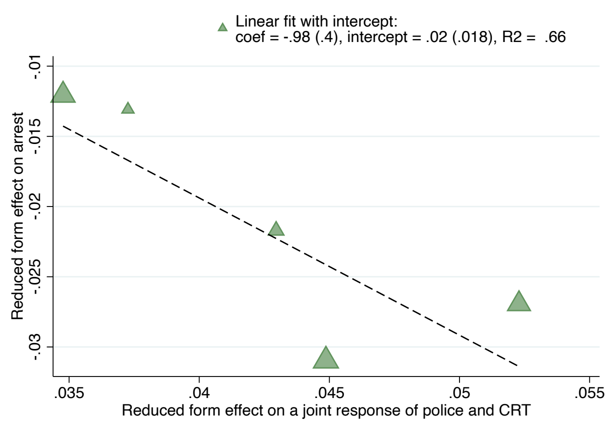
(a) Arrests vs. any CRT response



(b) Arrests vs. CRT-only responses



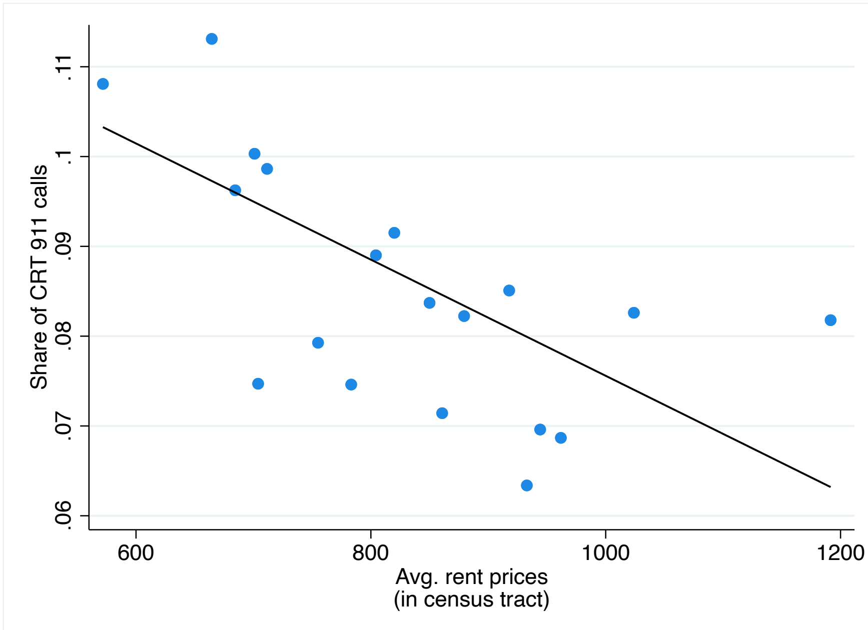
(c) Arrests vs. CRT-and-police responses



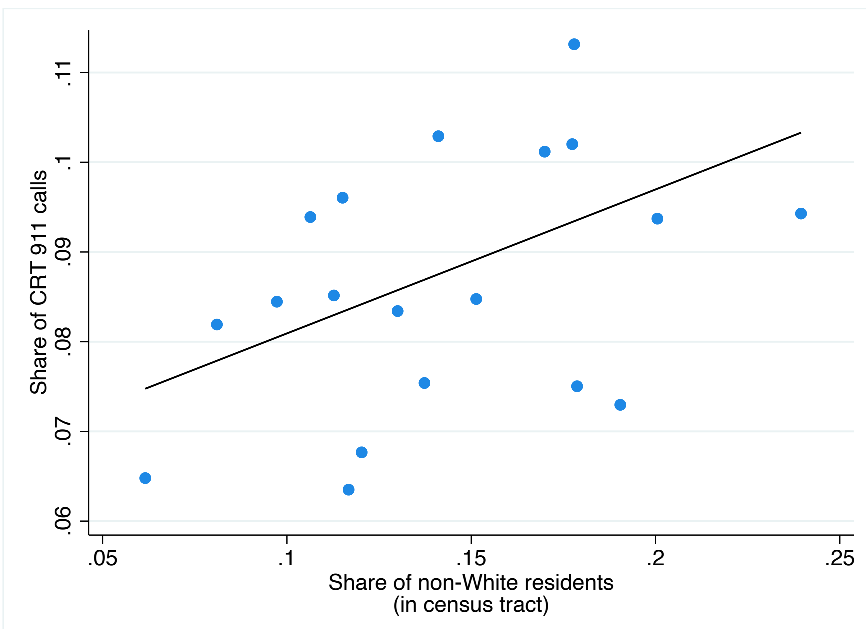
Notes: This figure shows the relationship between the reduced-form effect of each expansion on arrests (y-axis) and the expansion's impact on different response types (x-axis). Panel A reports effects on any CRT response. Panel B reports effects on CRT-only responses (with no police response). Panel C reports effects on responses in which both CRT and police are dispatched, not necessarily simultaneously. All regression lines are from weighted OLS based on the precision of the x-axis effects.

Figure A15: The correlation between CRT services and neighborhood characteristics

(a) CRT share of 911 calls and average rent



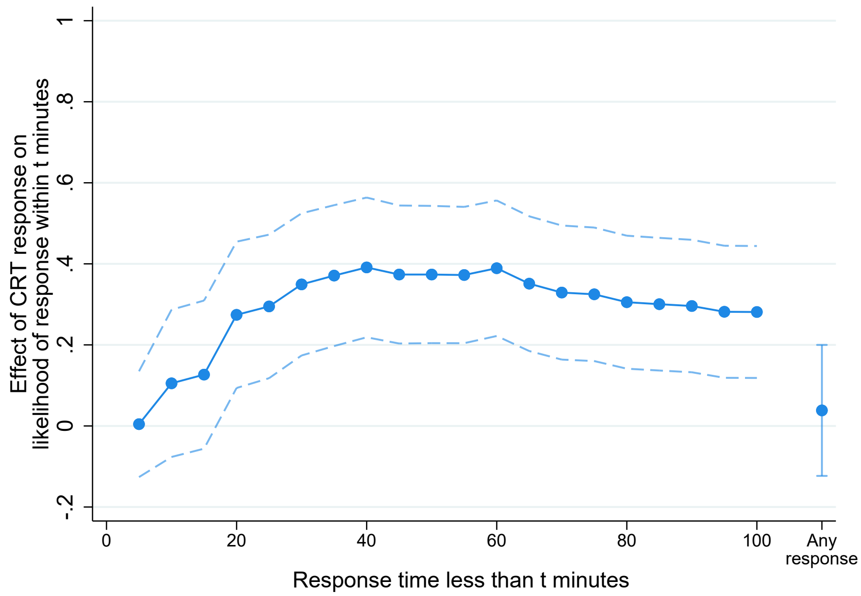
(b) CRT share of 911 calls and share of non-White residents



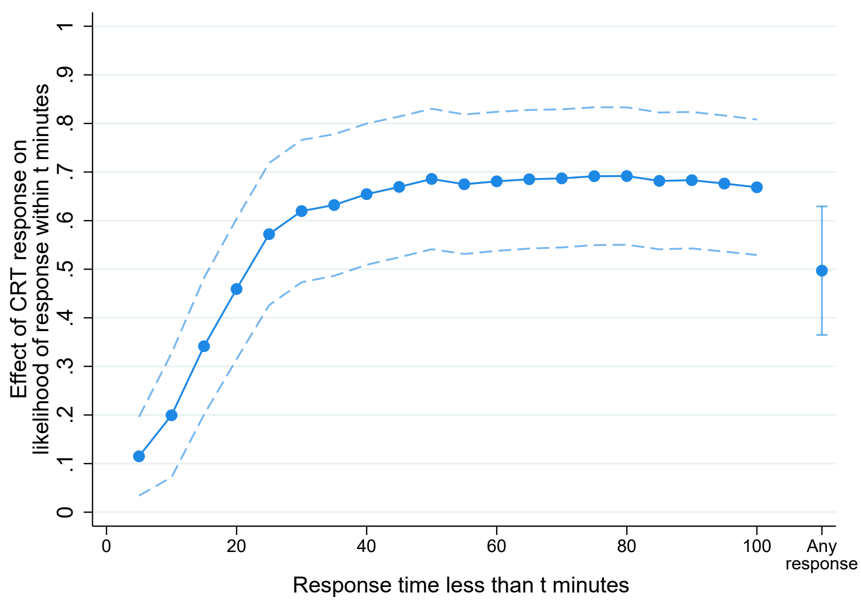
Notes: This figure shows correlations between CRT service use and neighborhood characteristics. Panel A plots the relationship between the CRT share of 911 calls and average rent prices. Panel B plots the relationship between the CRT share of 911 calls and the share of non-White residents. Neighborhood characteristics are measured using tract-level data from the American Community Survey and are merged to 911 call data based on the geographic coordinates of each 911 call.

Figure A16: Effect of a CRT response on likelihood of response within t minutes

(a) Expansions design



(b) Availability design



Notes: This figure shows the effect of a CRT response on the likelihood of a response within t minutes. Panel A reports estimates using the expansions design. Panel B reports estimates using the availability design. Each dot represents the coefficient from a 2SLS regression where the outcome is an indicator for a response occurring within t minutes.

A4 Appendix Tables

Table A1: Likelihood of police follow-up when a CRT arrives first

	Share of CRT-first calls
Police followup	0.035
with detox or mental health report	0.005
with crime report	0.005
with arrest	0.008
Calls where CRT arrives first	15,166
Total CRT calls	23,555

Notes: This table reports the frequency of police followup responses with various outcomes for calls where a CRT arrived first. Police followup is defined as police being dispatched after a CRT arrives on scene. Subcategories indicate the end result of call: detox or mental health hold, crime report filed, or arrest made.

Table A2: CRT expansions in Eugene and Springfield since 2014

	(1) Date	(2) Coverage	(3) Change
		Eugene	
Study baseline	As of 1/1/2014	11-3 a.m.	16
Expansion 1	11/1/2016	+ 7-10 a.m.	+4 (20 total)
Expansion 2	1/1/2017	+ 3-7 a.m	+4 (24 total)
		Springfield	
Study baseline	As of 1/1/2014	-	0
Expansion 3	1/14/2015	+ noon-11 p.m.	+11 (11 total)
Expansion 4	4/1/2017	+ 9-11 a.m. & 11 p.m.-2 a.m.	+7 (18 total)
Expansion 5	9/1/2017	+ 3-10 a.m.	+6 (24 total)

Notes: This table reports the start dates and resulting hour-of-day coverage of the five CRT expansions in Eugene and Springfield.

Table A3: Effect of CRTs on arrest resistance, use of force, and mortality (expansions instrument)

	Resisting arrest (1)	Use of force (2)	Either (3)	Dead on arrival (4)
CRT	0.006	0.002	0.007	-0.002
	(0.014)	(0.017)	(0.019)	(0.006)
Control complier mean	-0.002	0.009	0.006	0.004
Outcome mean	0.005	0.005	0.008	0.001
Number with outcome = 1	1,362	1,367	2,189	158
First-stage F-statistic	172.60	172.60	172.60	172.60
Observations	291,599	291,599	291,599	291,599

Notes: This table presents estimates of the impact of a CRT response on arrest resistance, use of force, either resisting or use of force, or being dead on arrival. We use a stacked design that combines all policy changes as instruments for a CRT response. Heteroskedasticity robust standard errors, clustered by call, are in parentheses. Significance stars indicate * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Robustness analysis of effects of CRTs using the expansion design

	Call responses		Call outcomes			
	Police (1)	Any response (2)	Arrest (3)	Invol. hold (4)	Detox (5)	Warrant (6)
Panel A: IV estimates of effect of CRT response						
CRT	-0.414***	0.038	-0.243***	-0.151***	-0.052***	0.022
	(0.084)	(0.082)	(0.062)	(0.023)	(0.013)	(0.026)
Control complier mean	0.962	0.962	0.321	0.173	0.055	-0.010
First-stage F-stat	172.59	172.59	172.59	172.59	172.59	172.59
Panel B: adding geographic tract controls						
CRT	-0.392***	0.062	-0.234***	-0.151***	-0.051***	0.023
	(0.083)	(0.081)	(0.062)	(0.023)	(0.013)	(0.026)
Control complier mean	0.938	0.938	0.312	0.173	0.054	-0.012
First-stage F-stat	175.35	175.35	175.35	175.35	175.35	175.35
Panel C: adding geographic tract and call nature controls						
CRT	-0.339***	0.118	-0.241***	-0.143***	-0.053***	-0.004
	(0.083)	(0.080)	(0.057)	(0.022)	(0.013)	(0.017)
Control complier mean	0.882	0.882	0.319	0.165	0.056	0.015
First-stage F-stat	165.45	165.45	165.45	165.45	165.45	165.45
Panel D: adding tract FEs						
CRT	-0.380***	0.075	-0.229***	-0.151***	-0.051***	0.025
	(0.083)	(0.081)	(0.062)	(0.023)	(0.013)	(0.026)
Control complier mean	0.925	0.925	0.306	0.173	0.054	-0.014
First-stage F-stat	172.95	172.95	172.95	172.95	172.95	172.95
Observations	291,599	291,599	291,599	291,599	291,599	291,599

Notes: This table explores the sensitivity of the estimates in Table 3 to the inclusion of additional controls. Each panel adds successively more controls. Robust standard errors, clustered by call, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Robustness of estimates to the sample of calls

		Nature-specific CRT likelihood		
	All	Any	CRT likely	CRT unlikely
	(1)	(2)	(3)	(4)
Panel A: First stage effect of CRT expansion on CRT response				
Post expansion	0.075***	0.077***	0.160***	0.040***
	(0.003)	(0.003)	(0.008)	(0.002)
Untreated mean	0.005	0.005	0.016	0.001
Panel B: IV estimates of CRT response (expansions design)				
CRT	-0.243***	-0.247***	-0.221***	-0.241*
	(0.062)	(0.061)	(0.046)	(0.145)
First-stage F-statistic	172.60	175.13	116.64	71.78
Observations	291,599	285,705	85,169	206,430
Panel C: First stage effect of CRT availability on CRT response				
Binary CRT availability	0.030***	0.030***	0.057***	0.015***
	(0.001)	(0.001)	(0.004)	(0.001)
Untreated mean	0.083	0.084	0.197	0.035
Panel D: IV estimates of CRT response (availability design)				
CRT	-0.045	-0.048	-0.065	0.057
	(0.041)	(0.041)	(0.043)	(0.102)
First-stage F-statistic	522.67	524.24	219.99	191.14
Observations	236,534	230,865	65,419	164,324

Notes: This table presents estimates from robustness analysis varying the sample of calls. Panels A and C look at the reduced form impact of both CRT expansions and the availability indicator. Panels B and D look at IV estimates of the impact of a CRT response on a call resulting in an arrest. Column 1 reports results based on our main sample. Columns 2-4 report results on different sample of calls which include : excluding call natures with a zero CRT response rate, a sub-set of the data with call nature types that are more likely to have a CRT respond ("CRT likely") and another sub-set of data with call nature types that are less likely to have CRT respond ("CRT unlikely"). Standard errors are in parentheses. Significance stars indicate * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Reduced form effects of each CRT expansion

	Call responses			Call outcomes			
	(1) CRT	(2) Police	(3) Any response	(4) Arrest	(5) Invol. hold	(6) Detox	(7) Warrant
Eugene 3-7 a.m., after Jan. 1 2017	0.11***	-0.076***	-0.014	-0.031	0.0011	-0.0024	-0.00058
	(0.011)	(0.028)	(0.028)	(0.023)	(0.0080)	(0.0047)	(0.011)
Eugene 7-10 a.m., after Nov. 1 2016	0.083***	-0.068**	-0.020	-0.012	0.0077	-0.0046	-0.0089
	(0.011)	(0.027)	(0.027)	(0.019)	(0.0069)	(0.0038)	(0.0085)
Springfield noon-11 p.m., after Jan. 14 2015	0.072***	-0.0071	0.028***	-0.013**	-0.015***	-0.0041***	0.0060***
	(0.0031)	(0.0083)	(0.0082)	(0.0059)	(0.0024)	(0.0012)	(0.0022)
Springfield 9-11 a.m. and 11 p.m.-2 a.m., after Apr. 1 2017	0.060***	-0.036***	-0.019**	-0.022***	-0.014***	-0.0038***	0.00038
	(0.0047)	(0.0095)	(0.0092)	(0.0073)	(0.0026)	(0.0015)	(0.0031)
Springfield 3-9 a.m., after Sept. 1 2017	0.10***	-0.078***	-0.025	-0.027**	-0.0057	-0.0035	-0.0073
	(0.010)	(0.019)	(0.019)	(0.013)	(0.0040)	(0.0026)	(0.0062)
Variance-weighted avg. effect	0.092	-0.064	-0.015	-0.023	-0.001	-0.004	-0.004
	(0.006)	(0.015)	(0.015)	(0.011)	(0.004)	(0.002)	(0.005)
Observations	291,599	291,599	291,599	291,599	291,599	291,599	291,599

Notes: This table reports the effect of CRT expansions on types of call responses and call outcomes. Each column reports the coefficients from a regression of the outcome variable on indicators for each CRT expansion. City-design and month-design fixed effects and controls for other policy changes are included. Standard errors clustered at the call level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Effects of CRTs on call response and arrest using the expansions design in pre-pandemic period

	Call responses			Call outcomes			
	CRT	Police	Any response	Arrest	Invol. hold	Detox	Warrant
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Reduced form effect of expansion							
Post expansion	0.075***	-0.023***	0.010	-0.016***	-0.011***	-0.003***	0.001
	(0.003)	(0.007)	(0.007)	(0.005)	(0.002)	(0.001)	(0.002)
Untreated mean	0.005	0.722	0.724	0.119	0.018	0.005	0.018
Panel B: IV estimates of effect of CRT response							
CRT		-0.308***	0.133	-0.210***	-0.145***	-0.046***	0.012
		(0.088)	(0.086)	(0.065)	(0.024)	(0.013)	(0.027)
Control complier mean		0.867	0.867	0.292	0.171	0.050	-0.000
First-stage F-stat		147.91	147.91	147.91	147.91	147.91	147.91
J test of overidentification							
χ^2 statistic		9.480	10.135	0.557	16.544	1.471	5.968
p-value		0.050	0.038	0.968	0.002	0.832	0.202
Hausman test							
χ^2 statistic		0.481	6.126	7.255	51.805	14.416	2.782
p-value		0.488	0.013	0.007	0.000	0.000	0.095
Panel C: OLS estimates of effect of CRT response							
CRT		-0.201***	0.372***	-0.035***	0.012***	0.002***	-0.012***
		(0.004)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)
Untreated mean		0.644	0.644	0.096	0.008	0.002	0.017
Observations		213,921	213,921	213,921	213,921	213,921	213,921

Notes: This table presents estimates of the impact of a CRT response on call responses and call outcomes using our expansions design. Panel A reports reduced form estimates of the effect of an expansion on these call outcomes. Panel B shows 2SLS estimates of the effect of a CRT response on call outcomes using indicators for each expansion as instruments. Panel C shows the association between a CRT response and these outcomes using the same set of controls. We use a stacked design that combines all policy changes as instruments for a CRT response. Heteroskedasticity robust standard errors, clustered by call, are in parentheses. Significance stars indicate * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Effects of CRTs on call response and arrest using the availability design in pre-pandemic period

	Call responses			Call outcomes			
	CRT (1)	Police (2)	Any response (3)	Arrest (4)	Invol. hold (5)	Detox (6)	Warrant (7)
Panel A: Reduced form effect of high CRT capacity							
High CRT capacity (=1)	0.029***	-0.007***	0.013***	-0.001	0.000	0.000**	-0.000
	(0.002)	(0.003)	(0.002)	(0.001)	(0.000)	(0.000)	(0.001)
Untreated mean	0.080	0.613	0.657	0.089	0.007	0.001	0.016
Panel B: IV estimates of effect of CRT response							
CRT		-0.242***	0.444***	-0.029	0.002	0.014*	-0.003
		(0.086)	(0.082)	(0.051)	(0.014)	(0.007)	(0.022)
Control complier mean		0.556	0.556	0.077	0.026	-0.007	0.004
First-stage F-stat		362.98	362.98	362.98	362.98	362.98	362.98
Hausman test							
χ^2 statistic		0.148	2.986	0.138	0.595	3.167	0.357
p-value		0.700	0.084	0.710	0.440	0.075	0.550
Equality with DinD LATE (p)		0.594	0.009	0.029	0.000	0.000	0.678
Panel C: OLS estimates of effect of CRT response							
CRT		-0.193***	0.382***	-0.030***	0.013***	0.002***	-0.011***
		(0.004)	(0.002)	(0.002)	(0.001)	(0.000)	(0.001)
Untreated mean		0.626	0.626	0.091	0.005	0.001	0.017
Observations		168,303	168,303	168,303	168,303	168,303	168,303

Notes: This table presents estimates of the impact of a CRT response on call responses and call outcomes using our availability design. Panel A reports reduced form estimates of the effect of an indicator for having high CRT capacity when the focal call comes on these call outcomes. Panel B shows 2SLS estimates of the effect of a CRT response on call outcomes using the CRT capacity indicator as an instrument for having a CRT response. Panel C shows the association between a CRT response and these outcomes using the same set of controls. Heteroskedasticity robust standard errors are in parentheses. Significance stars indicate * p < 0.10, ** p < 0.05, *** p < 0.01

Table A9: Sensitivity of estimated CRT effect on arrests to police busyness

	Diff-in-diff		Availability	
	(1)	(2)	(3)	(4)
CRT	-0.243***	-0.283***	-0.025	-0.045
	(0.062)	(0.064)	(0.041)	(0.041)
Control complier mean	0.321	0.361	0.061	0.080
First-stage F-stat	172.60	159.03	527.37	522.67
Observations	291,599	291,599	236,534	236,534
Control for police busyness	No	Yes	No	Yes

Notes: This table shows the sensitivity of the estimated impacts of dispatching CRTs on arrests to controlling for police busyness using the expansions and availability designs. Heteroskedasticity robust standard errors, clustered by call in columns (1) and (2), are in parentheses. Significance stars indicate * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A10: Robustness of effects to call volume

	Call responses		Call outcomes			
	Police	Any response	Arrest	Invol. hold	Detox	Warrant
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Baseline						
CRT	-0.414***	0.038	-0.243***	-0.151***	-0.052***	0.022
	(0.084)	(0.082)	(0.062)	(0.023)	(0.013)	(0.026)
Control complier mean	0.962	0.962	0.321	0.173	0.055	-0.010
First-stage F-stat	172.60	172.60	172.60	172.60	172.60	172.60
Panel B: Controlling for city × year FEs						
CRT	-0.780***	-0.314***	-0.337***	-0.103***	-0.046**	-0.031
	(0.120)	(0.118)	(0.088)	(0.032)	(0.018)	(0.039)
Control complier mean	1.314	1.314	0.418	0.132	0.049	0.043
First-stage F-stat	109.92	109.92	109.92	109.92	109.92	109.92
Panel C: Controlling for city-design specific year trends						
CRT	-0.440***	0.019	-0.231***	-0.129***	-0.037**	-0.016
	(0.102)	(0.099)	(0.074)	(0.027)	(0.015)	(0.032)
Control complier mean	0.981	0.981	0.316	0.161	0.042	0.026
First-stage F-stat	90.69	90.69	90.69	90.69	90.69	90.69
Panel D: Controlling for city-design-month number of calls						
CRT	-0.640***	-0.191**	-0.258***	-0.138***	-0.045***	0.002
	(0.097)	(0.095)	(0.070)	(0.025)	(0.014)	(0.031)
Control complier mean	1.191	1.191	0.337	0.167	0.049	0.008
First-stage F-stat	102.34	102.34	102.34	102.34	102.34	102.34
Observations	291,599	291,599	291,599	291,599	291,599	291,599

Notes: Significance stars indicate * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A11: Traditional diff-in-diff and Borusyak, Jaravel, and Spiess (2024)

	CRT (1)	Police (2)	Any response (3)	Arrest (4)	Invol. hold (5)	Detox (6)	Warrant (7)
Panel A: Traditional diff-in-diff							
Post expansion	0.075***	-0.031***	0.003	-0.018***	-0.011***	-0.004***	0.002
	(0.003)	(0.006)	(0.006)	(0.005)	(0.002)	(0.001)	(0.002)
Dependent mean	0.082	0.603	0.654	0.086	0.007	0.002	0.016
Observations	291,599	291,599	291,599	291,599	291,599	291,599	291,599
Panel B: BJS imputation estimator							
Post expansion	0.076***	-0.037***	0.001	-0.023***	-0.008**	-0.003*	0.002
	(0.003)	(0.011)	(0.011)	(0.008)	(0.003)	(0.002)	(0.004)
Dependent mean	0.005	0.725	0.727	0.120	0.019	0.005	0.018
Observations	38,661	38,661	38,661	38,661	38,661	38,661	38,661

Notes: This table compares the traditional difference in differences to the imputation approach of Borusyak, Jaravel, and Spiess (2024). The traditional diff-in-diff differs only from our main reduced form approach by using an indicator for any expansion rather than separate indicators for each expansion; it therefore includes call hour-city-design and call month-design FEs. The BJS specification uses only the ever-treated observations, and includes call hour and call month FEs. To match the implicit weights in our main 2SLS specification, all reduced-form regressions are variance reweighted at the design level by the ratio of the variance of the residualized first stages for CRT using each expansion and an indicator for any expansion, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: 2SLS estimates of CRT effects using the availability design

	Call responses		Call outcomes			
	Police (1)	Any response (2)	Arrest (3)	Invol. hold (4)	Detox (5)	Warrant (6)
Panel A: IV estimates of effect of CRT response						
CRT	-0.228***	0.497***	-0.045	-0.004	0.012**	-0.020
	(0.071)	(0.067)	(0.041)	(0.011)	(0.005)	(0.018)
Control complier mean	0.412	0.412	0.061	0.025	-0.006	0.011
First-stage F-stat	522.67	522.67	522.67	522.67	522.67	522.67
Panel B: adding geographic tract controls						
CRT	-0.237***	0.488***	-0.048	-0.005	0.012**	-0.020
	(0.071)	(0.067)	(0.041)	(0.011)	(0.005)	(0.018)
Control complier mean	0.421	0.421	0.063	0.025	-0.006	0.012
First-stage F-stat	518.57	518.57	518.57	518.57	518.57	518.57
Panel C: adding geographic tract and call nature controls						
CRT	-0.239***	0.484***	-0.051	-0.009	0.013**	-0.019
	(0.072)	(0.069)	(0.044)	(0.012)	(0.006)	(0.018)
Control complier mean	0.424	0.424	0.069	0.031	-0.006	0.012
First-stage F-stat	531.27	531.27	531.27	531.27	531.27	531.27
Panel D: adding tract FEs						
CRT	-0.237***	0.487***	-0.049	-0.005	0.012**	-0.021
	(0.071)	(0.067)	(0.041)	(0.011)	(0.005)	(0.018)
Control complier mean	0.422	0.422	0.065	0.025	-0.005	0.013
First-stage F-stat	516.72	516.72	516.72	516.72	516.72	516.72
Observations	236,534	236,534	236,534	236,534	236,534	236,534

Notes: This table explores the sensitivity of the estimates in Table 4 to the inclusion of additional controls. Each panel successively adds more controls. Robust standard errors in parentheses. Significance stars indicate * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A13: Reduced form effects on reported incidents

		Non-criminal		Criminal			
	Any report	Mental health	Overdose	Any	Violent	Drug	Property
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Expansions design							
Post expansion	-0.012**	-0.01***	-0.00288**	-0.00756	-0.00267	-0.00244**	-0.00165
	(0.006)	(0.002)	(0.001)	(0.005)	(0.003)	(0.001)	(0.003)
Untreated mean	0.225	0.021	0.007	0.149	0.029	0.006	0.045
Observations	291,599	291,599	291,599	291,599	291,599	291,599	291,599
Panel B: Availability design							
Binary CRT availability	-0.00165	-0.0000305	0.000147	-0.00189	-0.000377	0.000408*	0.0000608
	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
Untreated mean	0.167	0.008	0.003	0.125	0.029	0.003	0.029
Observations	236,534	236,534	236,534	236,534	236,534	236,534	236,534

Notes: This table presents reduced form estimates from regressions of report outcomes on indicators for CRT expansions or the availability indicator. Panel A shows reduced form estimates using the expansions design, while Panel B shows estimates from the availability design. Heteroskedasticity robust standard errors, clustered by call, are in parentheses. Significance stars indicate * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A14: Effects of arrival timing for calls with both police and a CRT response

	(1) Pred. arrest	(2) Arrest	(3) Invol. hold	(4) Detox	(5) Warrant
CRT arrives 2+ minutes before police	0.000704	-0.0292	-0.0207	-0.00937	0.0143
	(0.00192)	(0.0394)	(0.0239)	(0.00667)	(0.0164)
Police arrives 2+ minutes before CRT	0.000341	-0.0169	-0.0000639	-0.00446	0.00331
	(0.00102)	(0.0268)	(0.0200)	(0.00804)	(0.00797)
Early police = early CRT (p)	0.838	0.714	0.260	0.436	0.511
Dependent mean for same-time arrival	0.077	0.106	0.053	0.010	0.005
Observations	704	704	704	704	704

Notes: This table examines whether the relative arrival order of police and CRTs affects arrest outcomes among calls in which both units are dispatched. The sample is restricted to calls where police and CRT are dispatched within two minutes of each other and the final dispatch occurs at least five minutes before the first unit arrival. Coefficients come from a regression of arrest outcomes on indicators for the police arriving at least two minutes before a CRT and for a CRT arriving at least two minutes before police. City-by-year and hour-of-day fixed effects are included. Standard errors are clustered at call level.

References

Chandon Adger, Matthew Ross, and CarlyWill Sloan. The effect of field training officers on police use of force. Working paper, 2025.

Joshua D. Angrist. Lifetime earnings and the Vietnam Era draft lottery: Evidence from social security administrative records. *The American Economic Review*, 80(3):313–336, 1990.

Bocar Ba, Meghna Baskar, Tony Cheng, and Rei Mariman. Understanding demand for police alternatives. Working Paper, 2024.

Bocar A Ba, Patrick J. Bayer, Nayoung Rim, Roman Rivera, and Modibo Sidibé. Police officer assignment and neighborhood crime. National Bureau of Economic Research Working Paper No. 29243, 2021a.

Bocar A Ba, Dean Knox, Jonathan Mummolo, and Roman Rivera. The role of officer race and gender in police-civilian interactions in chicago. *Science*, 371(6530):696–702, 2021b.

Bocar A. Ba, Tony Chen, Patton Cheng, Martha C. Eies, and Justin E. Holz. What is the best response? examining the impact of police and their alternatives. *NBER Working Paper Series*, (34344), 2025.

Etienne Blais, Marjolaine Landry, Nicolas Elazhary, Sebastian Carrier, and Anne-Marie Savard. Assessing the capability of a co-responding police-mental health program to connect emotionally disturbed people with community resources and decrease police use-of-force. *Journal of Experimental Criminology*, 18:41–65, 2022.

Jordi Blanes i Vidal and Tom Kirchmaier. The effect of police response time on crime clearance rates. *The Review of Economic Studies*, 85(2):855–891, 2018.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, page rdae007, 2024.

Anthony A Braga, John M MacDonald, and James McCabe. Body-worn cameras, lawful police stops, and nypd officer compliance: A cluster randomized controlled trial. *Criminology*, 60(1):124–158, 2022.

Aaron Chalfin and Justin McCrary. Are us cities underpoliced? theory and evidence. *Review of Economics and Statistics*, 100(1):167–186, 2018.

Aaron Chalfin, Benjamin Hansen, Emily K Weisburst, and Morgan C Williams Jr. Police force size and civilian race. *American Economic Review: Insights*, 4(2):139–158, 2022.

Galia Cohen and Matt Bagwell. The state of mental health training in basic police curricula: A national level examination. *Journal of Public Affairs Education*, 29(3):327–349, 2023.

Thomas S. Dee and Jaymes Pyne. A community response approach to mental health and substance abuse crises reduced crime. *Science Advances*, 8(23):1–9, 2022. doi: 10.1126/sciadv.abm2106. URL https://www.science.org/doi/abs/10.1126/sciadv.abm2106.

Manasi Deshpande and Michael Mueller-Smith. Does welfare prevent crime? the criminal justice outcomes of youth removed from ssi. *The Quarterly Journal of Economics*, 137(4): 2263–2307, 2022.

Rafael Di Tella and Ernesto Schargrodsky. Do police reduce crime? estimates using the allocation of police forces after a terrorist attack. *American Economic Review*, 94(1): 115–133, 2004.

DOJ. Final Report of the President's Task Force on 21st Century Policing. *Department of Justice*, May 2015. Washington, DC: Office of Community Oriented Policing Services.

Mirko Draca, Stephen Machin, and Robert Witt. Panic on the streets of london: Police, crime, and the july 2005 terror attacks. *American Economic Review*, 101(5):2157–81, 2011.

Oeindrila Dube, Sandy Jo MacArthur, and Anuj K Shah. A cognitive view of policing. *The Quarterly Journal of Economics*, 140(1):745–791, 2025.

Eugene Police. Policy c4.10: Cahoots, 2017. Available at https://lawenforcementactionpartnership.org/wp-content/uploads/2023/06/CAHOOTS-procedure.pdf.

William N Evans and Emily G Owens. Cops and crime. *Journal of Public Economics*, 91 (1-2):181–201, 2007.

Gino Fanelli. City Council approves consulting contract with CAHOOTS operator, 2020. Available at https://www.wxxinews.org/local-news/2020-12-16/city-council-approves-consulting-contract-with-cahoots-operator.

Amy Finkelstein and Nathaniel Hendren. Welfare analysis meets causal inference. *Journal of Economic Perspectives*, 34(4):146–167, 2020.

Lisa Rose Gagnon, Liz Reetz, and Sarah Radcliffe. Mental Health Holds and Civil Commitment, 2022. Available at https://www.droregon.org/disability-rights-oregon-resources/mental-health-holds-and-civil-commitment.

Robert Gonzalez and Sarah Komisarow. Community monitoring and crime: Evidence from chicago's safe passage program. *Journal of Public Economics*, 191, 2020.

James J Heckman and Rodrigo Pinto. Unordered monotonicity. *Econometrica*, 86(1):1–35, 2018.

Margaret Heslin, Lynne Callaghan, Martin Packwood, Vincent Badu, and Sarah Byford. Decision analytic model exploring the cost and cost-offset implications of street triage. *BMJ Open*, 6, 2016.

Mark Hoekstra and CarlyWill Sloan. Does race matter for police use of force? evidence from 911 calls. *American Economic Review*, 112:827–860, 2022.

Harry J. Holzer, Lawrence F. Katz, and Alan B. Krueger. Job queues and wages. *The Quarterly Journal of Economics*, 103(4):773–807, 1988.

Amos Irwin and Betsy Pearl. The community responder model: How cities can send the right responder to every 911 call, 2020. Available at https://www.americanprogress.org/article/community-responder-model/.

Oliver Jenkins, Stephen Dye, Franklin Obeng-Asare, Nam Nguyen, and Nicola Wright. Police liaison and section 136: comparison of two different approaches. *BJPsych Bull*, 41:76–82, 2017.

Zeynal Karaca and Brian J. Moore. Costs of emergency department visits for mental and substance use disorders in the united states, 2017. Hcup statistical brief #257, Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality, Rockville, MD, 2020. URL https://hcup-us.ahrq.gov/reports/statbriefs/sb257-ED-Costs-Mental-Substance-Use-Disorders-2017.jsp. Revised October 2020.

Patrick Keown, Jo French, Graham Gibson, Eddy Newton, Steve Cull, Paul Brown, Jo Parry, Diana Lyons, and Iain McKinnon. Too much detention? street triage and detentions under section 136 mental health act in the north-east of england: a descriptive study of the effects of a street triage intervention. *BMJ Open*, 6, 2016.

Ashley Krider, Regina Huerter, Kirby Gaherty, and Andrew Moore. Responding to individuals in behavioral health crisis via co-responder models: The roles of cities, counties, law enforcement, and providers, 2020. Available at https://www.prainc.com/wp-content/uploads/2020/03/RespondingtoBHCrisisviaCRModels.pdf.

Steven Mello. More cops, less crime. *Journal of Public Economics*, 172:174–200, 2019.

Steven Mello. Fines and financial wellbeing. *Review of Economic Studies*, 2024.

Amalia R Miller and Carmit Segal. Do female officers improve law enforcement quality? effects on crime reporting and domestic violence. *The Review of Economic Studies*, 86(5): 2220–2247, 2019.

Emily Owens, David Weisburd, Karen L Amendola, and Geoffrey P Alpert. Can you build a better cop? experimental evidence on supervision, training, and policing in the community. *Criminology & Public Policy*, 17(1):41–87, 2018.

Deepak Premkumar. Public scrutiny, police behavior, and crime consequences: Evidence from high-profile police killings. 2019.

Roman Rivera. Are 'bad' cops better police? the trade-off between officer aggression and public safety. Working paper, 2025a.

Roman Rivera. Do peers matter in the police academy? *American Economic Journal: Applied Economics*, 17(2):127–164, April 2025b. doi: 10.1257/app.20220348. URL https://www.aeaweb.org/articles?id=10.1257/app.20220348.

Michael S Rogers, Dale E McNiel, and Renée L Binder. Effectiveness of police crisis intervention training programs. *The Journal of the American Academy of Psychiatry and the Law*, 47(4):414–421, 2019.

Jonathan Tebes and Jeffrey Fagan. Do pedestrian stops deter crime? evidence from reforming "stop and frisk". 2025. Working paper.

Alan Torres. NYC mayor candidate shouts out CAHOOTS in Eugene as program to emulate, 2025. Available at https://www.registerguard.com/story/news/politics/2025/10/17/mamdani-praises-cahoots-and-eugene-in-nyc-mayor-debate/86753895007/.

U.S. Senate. CAHOOTS Act, S.764, 117th Congress (2021-2022), 2021. URL https://www.congress.gov/bill/117th-congress/senate-bill/764.

Susan Walker, Euan Mackay, Phoebe Barnett, Luke Sheridan Rains, Monica Leverton, Christian Dalton-Locke, Kylee Trevillion, Brynmor Lloyd-Evans, and Sonia Johnson. Clinical and social factors associated with increased risk for involuntary psychiatric hospitalisation: a systematic review, meta-analysis, and narrative synthesis. *The Lancet Psychiatry*, 6(12):1039–1053, 2019.

Amy C Watson and Michael T Compton. What research on crisis intervention teams tells us and what we need to ask. *The Journal of the American Academy of Psychiatry and the Law*, 47(4):422–426, 2019.

Emily K Weisburst. Safety in police numbers: Evidence of police effectiveness from federal cops grant applications. *American Law and Economics Review*, 21:81–109, 2019.

White Bird Clinic. *CAHOOTS Media Guide*, 2020. URL https://whitebirdclinic.org/wp-content/uploads/2020/06/CAHOOTS-Media-Guide-20200626.pdf. Accessed April 1, 2025.

Morgan C Williams, Nathan Weil, Elizabeth A Rasich, Jens Ludwig, Hye Chang, and Sophia Egrari. Body-worn cameras in policing: Benefits and costs. 2021.