What is the Best Response? Examining the Impact of Police and Their Alternatives [*]

Bocar A. Ba[†] Patton Chen [‡] Tony Cheng[§] Martha Eies [¶] Justin Holz [||]

Please do not circulate or cite without authorization

March 20, 2025

Abstract

Recent protests over police behavior have led cities to implement civilian crisis response programs as alternatives to traditional policing. However, little is known about their impacts and cost-effectiveness. This paper evaluates Durham, North Carolina's, HEART program, a civilian-led crisis response initiative designed to handle nonviolent 911 calls and minimize police intervention. Using a difference-in-differences approach, we find that HEART significantly reduces crime reports, arrests, and response times, with the effects driven by remote and in-person civilian responses, while co-response teams have minimal impact. Remote response initially increases backup needs but overall HEART fosters trust in 911, leading to greater engagement for lower-priority incidents. To assess cost-effectiveness, we design a contingent valuation survey to estimate public willingness to pay (WTP). Our results indicate strong public support, with WTP significantly surpassing the program's net cost. Applying the Marginal Value of Public Funds framework, we find that HEART's fiscal savings fully offset its direct costs, making it a self-sustaining intervention.

[*]Thanks to seminar participants at GSU and Penn LDI. Any errors are our own.

[†]Duke and NBER. Email: bocar.ba@duke.edu.
[‡]Duke. Email: bingjun.chen@duke.edu.
[§]Duke. Email: tony.cheng@duke.edu.
[¶]Duke. Email: martha.eies@duke.edu.
[||]Michigan. Email: holzj@umich.edu.

1 Introduction

For decades, police have responded to nearly all 911 calls, including those related to nonviolent incidents such as mental health crises, homelessness, and substance use. In response to concerns about overpolicing, cities have introduced civilian-led crisis response programs to handle such emergencies without law enforcement (Kaba et al., 2021; Akbar, 2020; Phelps, 2024). These programs aim to deescalate crises and reduce unnecessary arrests, but their impact remains largely untested. Do they improve public safety? Which models work best? For whom? Are they cost-effective?

Figure 1 highlights the sharp rise in nonpolice crisis response programs. After the launch of CAHOOTS in Eugene, Oregon, in 1989, growth remained minimal for decades, only accelerating post-2020 following widespread protests against police violence. Despite this expansion, rigorous evaluations of these programs remain limited. As cities scale these models, assessing their effects on enforcement, public safety, and costs is critical. A key question is whether crisis response should remain within policing or be outsourced (Hart et al., 1997).We examine this in crisis response, where shifting from police to civilian teams involves quality-cost trade-offs, especially when de-escalation success and community trust are hard to formalize in contracts.[1]

This paper evaluates Durham's HEART program, which deploys remote, in-person civilian and co-response teams that pair a police officer with a civilian. We find that relative to the outcomes of HEART-ineligible cases, crime reports, arrests, and response times are lower for calls that receive a HEART response. These effects are driven by in-person teams, while the co-response teams have minimal impact. Remote response initially increases backup needs but builds trust in 911, encouraging calls for lower-priority incidents. To assess the program's value, we design a contingent valuation survey—a novel approach for crisis response programs—and find strong public support. Our estimates show that HEART's fiscal savings exceed its direct costs, indicating that the program is self-sustaining.

Alternative crisis response models face key challenges: Call urgency varies, making it unclear when civilian teams can replace police; enforcement reductions may signal true safety gains or shifts in reporting; and effectiveness depends on whether the responses are on-scene, are remote, or involve police. Most US programs do not clearly distinguish between these response models, but HEART's centralized 911 system allows direct comparison. Beyond measuring effectiveness, policymakers must also assess costs and benefits to compare across jurisdictions and program types. A systematic approach is needed to determine whether these programs provide sufficient value to justify their expansion.

We address these challenges in several steps. First, we develop a framework to clarify the trade-offs among variations of police alternatives, whether remote, on-scene, civilian-led or police-

[1]Similar trade-offs arise in other topics related to other applications in the legal system, e.g, prison management (Mukherjee, 2021; Mukherjee and Sanders, 2021; Yegen, 2021), highlighting broader concerns about cost, accountability, and service quality.

assisted, as the impacts of each are theoretically ambiguous. Effective crisis response depends on appropriate matching of interventions to incident severity, responder costs, and social outcomes. Remote teams are best suited for low-risk calls, in-person teams for moderate crises, and co-response teams for high-risk situations. However, differences in costs and evolving response thresholds create uncertainty. This highlights the need for empirical evaluation to determine whether a shift in emergency response from police to alternative models improves public welfare.

Second, to evaluate HEART's impact, we use a difference-in-differences (DiD) approach, comparing cases that were eligible and ineligible for a HEART response before and after the program's launch. We determine eligibility by linking 911 call data to response records. Cases deemed ineligible, primarily those involving suspicious activity, remain under traditional police response protocols and exhibit similar pretrends in outcomes to the eligible cases. This design allows us to isolate the effects of crisis teams, phone-based interventions, and co-response on public safety outcomes. HEART responses fall into three categories: Crisis Call Diversion, a remote civilian response for mental health crises; the Community Response Teams, an on-scene civilian-led response for welfare checks and behavioral health issues; and Co-Response, whereby a police-clinician team handles higher-risk incidents. In this paper, we will refer to Crisis Call Diversion as "remote civilian response," Community Response Team as "on-scene civilian response," and Co-Response as "co-response" to ensure clarity and consistency in terminology.

We begin our analysis by examining how HEART and police responses impact public safety using a DiD approach. HEART's introduction significantly reduced crime reports by 9.2 percentage points (pp) and arrests by 2.8 pp (a 56% drop from a 5% baseline), with effects stabilizing over time. Police-handled cases showed no meaningful changes, confirming that HEART's impact stems from direct intervention. Response times fell by 5.9 minutes, while dispatch times dropped 37%. Backup requests initially doubled before stabilizing. Reductions were the largest for nonviolent offenses, and HEART lowered violent crime reports and arrests more than police. Overall, HEART reduced enforcement without compromising safety, though initial backup needs were higher.

HEART's long-term impact on 911 engagement sheds light on whether alternative crisis response fosters trust in emergency services or discourages future help-seeking. In contrast to what we observe for police responses after high-profile incidents, where trust is often eroded (Desmond et al., 2016; Rivera and Ba, 2023; Ang et al., 2024) or gun violence incident (Mikdash and Zaiour, 2024), HEART does not reduce future 911 calls. We see no significant short-run effect (1—31 days after an incident) on the likelihood of a repeat call for either HEART or police responses, suggesting comparable crisis resolution. In the longer term (beyond 31 days), both HEART and police responses increase future call probability (by 4.1 and 3.3 pp, respectively), suggesting that HEART fosters continued engagement rather than discouraging reliance on emergency services.

HEART's effectiveness varies by call type, reflecting differences across incidents in enforcement needs and deescalation potential. Mental health crises and disturbances are well suited for civilian teams, while violent or domestic incidents often require police involvement. HEART consistently re-

duces crime reporting, particularly for lower-risk calls such as trespassing and nuisance complaints, while police responses increase documentation, especially for violent or ambiguous incidents. For arrests, we see a similar pattern: HEART significantly reduces arrests for minor offenses, whereas police responses leave arrest rates unchanged. However, HEART teams frequently request backup, particularly for mental health and disturbance calls, highlighting a key operational trade-off. While the program minimizes law enforcement contact, it requires greater coordination and resources to ensure effective crisis management.

HEART's impact on arrest patterns further shows how alternative crisis response models shape law enforcement interactions, particularly for groups disproportionately affected by traditional policing. Pre-HEART arrest rates were highest among Black individuals, men, and people aged 25–39. The DiD estimates show that HEART significantly reduces arrests for these groups while the effects for Hispanic, White, female, and younger individuals are smaller and statistically insignificant. The largest declines occur among Black individuals and men (-0.024, $p < 0.01$ and -0.021, $p < 0.01$), suggesting HEART may help mitigate disparities in policing outcomes. The program's effect on 25- to 39-year-olds, a group overrepresented in arrest statistics, further supports its potential to reduce unnecessary law enforcement contact. Even among HEART-eligible cases handled by police, arrest rates remain highest for Black individuals and men, reinforcing persistent disparities. The stark contrast in outcomes of police and HEART responses suggests that these shifts are driven by changes in enforcement practices rather than differences in case selection, underscoring the potential of nonpolice crisis response models to reduce law enforcement interactions in communities historically subject to overpolicing (Derenoncourt, 2022; Cox et al., 2024).

We decompose HEART's effects by response model: remote civilian responses, on-scene civilian responses, and co-response teams, where police are paired with a civilian. This distinction allows us to compare fully civilian interventions with those that integrate law enforcement. Using a DiD approach, we compare the outcomes after each type of HEART intervention to the outcomes of ineligible cases. Both remote and in-person responses significantly reduce crime reports and arrests, albeit by different magnitudes. Remote response lowers crime reports by 3.5 pp ($SE = 0.010$), while in-person response achieves a larger reduction of 7.3 pp ($SE = 0.004$). Similarly, in-person response has the strongest effect on arrests, reducing them by 1.9 pp ($SE = 0.002$), while remote response slightly increases arrests by 1.3 percentage points ($SE = 0.003$). The co-response has minimal and statistically insignificant effects on crime reports and arrests, implying that when police are present, enforcement patterns largely mirror those under traditional policing.

Backup trends highlight early challenges. Remote response initially required significant backup support, which increased by 7.6 pp ($SE = 0.018$), while for in-person response, there was a smaller increase of 3.2 pp ($SE = 0.011$). These patterns suggest an adjustment period, particularly for remote interventions with no on-scene presence. Over time, backup needs declined, indicating improved triage and crisis management. The differences in the impacts on response time underscore these distinctions: Remote response significantly reduces the total response time by 25.9 minutes

($SE = 2.44$), while a response by in-person teams increases the time on scene by 5.7 minutes ($SE = 1.02$) and co-response extends the response duration by 4.7 minutes ($SE = 1.55$).

To assess whether HEART prevents crisis escalation, we examine its impact on police use of force (Ba et al., 2021; Hoekstra and Sloan, 2022) and future violent crime–related calls. Relative to police-handled cases, remote and in-person responses reduce use-of-force incidents, with the latter showing the largest decline (-2.1, $SE = 0.006$). Remote response exhibits more variability in these outcomes (-1.2, $SE = 0.009$), likely because of the early reliance on backup. Co-response has little effect (-0.3, $SE = 0.007$): When police are involved, use-of-force rates remain unchanged. Similarly, in-person response significantly reduces future violent crime–related calls (-1.8, $SE = 0.005$), while for remote response, the estimates are noisier, and co-response again has little impact. Moreover, HEART's expansion reshaped 911 call composition. Remote and in-person responses significantly lowered call priority, suggesting increased public confidence in seeking assistance for lower-risk incidents. Co-response and police-only responses showed smaller declines. We conclude from this that traditional law enforcement remains associated with more urgent cases. The findings suggest that HEART not only reduces enforcement but also builds trust in emergency services, leading to greater engagement for nonemergency support.

Next, we evaluate the welfare effects of HEART using the Marginal Value of Public Funds (MVPF) framework (Hendren, 2016; Hendren and Sprung-Keyser, 2020), which quantifies the benefits individuals receive per dollar of net government spending. The MVPF framework incorporates three primary components: the benefits of the policy, the direct costs of the program, and the fiscal externalities generated by the program. HEART's direct cost per response is $1,191 (City of Durham, 2023). There are two primary sources of fiscal externalities of HEART, which we analyze using our DiD approach. The first is the impact on the emergency response costs and the second is changes in the outcomes experienced by those served by HEART.

Police responses cost $67.53 per hour, while alternative models vary: remote crisis response is the least expensive ($32.50/hour), in-person civilian response is the most expensive ($98.09/hour), and co-response falls in between ($73.90/hour). Remote response yields the largest savings, reducing costs by $48.90 per call, while in-person response costs $30.91 more. Police-only responses see a modest $3.41 per-call savings, totaling $352,103. While remote response is cost-effective, its limited scale constrains budget impact. In-person response, though costlier, may offer long-term savings by reducing arrests and incarceration, highlighting a trade-off between cost-efficiency and public safety benefits. We also use our treatment effect estimates to understand the fiscal externalities from HEART. we find that fiscal savings from reduced victimization, investigations, and incarceration total $2,300 per response.

Since these savings more than offset operational expenses, HEART effectively funds itself through fiscal externalities, suggesting that the program fully covers its costs and the MVPF is infinity as long as the willingness to pay for the program is positive. Evaluating the willingness to pay for alternative crisis response models requires assessing not only their effectiveness in reducing crime (the use

value) but also their broader societal value (the non-use value). While prior studies have examined cost savings and enforcement outcomes, they often overlook a key question: how much do communities value these programs? Without a measure of willingness to pay (WTP), it is difficult to determine whether HEART's perceived benefits justify its costs from the perspective of both direct beneficiaries and the broader public.

We address this question by estimating the WTP for HEART using a contingent valuation (CV) survey, a widely used method in public finance for valuing environmental and criminal justice nonmarket goods (Chalfin, 2015; Bishop et al., 2017; Domínguez and Scartascini, 2024). Our pilot survey asks a representative sample of the US how much they would be willing to pay in taxes to support HEART, ensuring responses reflect both direct and indirect benefits. We estimate an average WTP of $67.50 per resident per year, or $5,872 per HEART response. Notably, 73% of respondents report a WTP that exceeds HEART's highest per-resident cost of $20.54 in 2024. Given the substantial WTP, the program yields an MVPF of infinity, indicating that its economic benefits fully cover its costs. These findings highlight both strong public support and financial sustainability, positioning HEART as a cost-effective alternative to traditional police-led crisis response.

All together, we find that HEART's direct cost per response is $1,191, while fiscal savings from reduced victimization, investigations, and incarceration total $2,901 per response. Morever, the willingness to pay for the program is large and exceeds the direct cost of the program, even without considering the savings from fiscal externalities. Since these savings more than offset operational expenses, HEART effectively funds itself through fiscal externalities. Given the substantial WTP, the program yields an MVPF of infinity, indicating that its economic benefits fully cover its costs. These findings highlight both strong public support and financial sustainability, positioning HEART as a cost-effective alternative to traditional police-led crisis response.

Literature Review Our study contributes to research on public safety, crisis response, and resource allocation. While increased police presence can reduce crime (Vidal and Kirchmaier, 2018; Chen et al., 2023),[2] negative interactions may erode trust and deter reporting (Desmond et al., 2016; Rivera and Ba, 2023; Ang et al., 2024). De-escalation training shows promise (Owens et al., 2018; Roth and Sant'Anna, 2023; Dube et al., 2024; Ross and Sloan, 2024), yet the effectiveness of police as first responders in mental health crises, substance use, and domestic disturbances remains uncertain (Ba et al., 2024; Foureaux Koppensteiner et al., 2024). Alternative models, such as virtual crisis response (Gates et al., 2024), show potential, but evaluations of civilian-led interventions in the U.S. are limited. Our study examines whether shifting emergency response from police to civilian crisis teams improves outcomes, building on evidence that specialized responders reduce enforcement while enhancing crisis resolution (Dee and Pyne, 2022; Deza et al., 2023; Davis et al., 2024). We also contribute to the literature on public resource allocation and police funding, includ-

[2]For an extensive review, see (Chalfin and McCrary, 2017).

ing research on officer hiring (Evans and Owens, 2007; Mello, 2019), increased law enforcement budgets (Bursztyn et al., 2023; Ba et al., 2023b,a; Moreno-Medina et al., 2025), and the broader effects of funding distribution on communities (Derenoncourt, 2022; Chyn et al., 2022; Cox et al., 2025). By assessing both effectiveness and cost-efficiency, our findings inform policy discussions on optimizing emergency response and resource allocation.

Second, our paper contributes to the literature on the welfare effects of public policies (Kuziemko et al., 2015; Hendren and Sprung-Keyser, 2020; Metcalfe and Roth, 2025) by providing one of the first estimates of public demand for police alternatives. While prior work has examined the welfare impacts of policing, victimization, and labor market outcomes (Chalfin and McCrary, 2018; Stevenson and Mayson, 2022; Cullen et al., 2022), we extend this research by directly measuring public preferences for alternative safety policies. Using a contingent valuation approach with a multiple price list (Anderson et al., 2007; Allcott and Kessler, 2019), we estimate WTP and assess trade-offs between traditional policing and alternatives. This method captures both use-value and non-use value (Chalfin, 2015), offering a more comprehensive view of public preferences. Additionally, we evaluate the fiscal externalities of police alternatives, providing a framework to quantify their broader economic impact. Our findings inform cost-benefit analyses in other policy settings (Deshpande and Mueller-Smith, 2022), helping policymakers assess the direct and indirect welfare effects of reallocating resources from policing to alternative public safety measures.

Plan The remainder of the paper is organized as follows. The next section provides background on emergency response in the US and Durham's HEART program. Section 3 describes the data and key descriptive statistics. Section 4 details the empirical strategy. Sections 5 and 6 examine the impact of HEART and the different response models on public safety, trust, and labor costs. Section **??** evaluates the program's cost-effectiveness using a contingent valuation survey and the MVPF framework. The final section concludes.

2 Background on Emergency Responses

2.1 Crisis Response Programs in the US

The 2020 murder of George Floyd accelerated efforts to rethink public safety (Kaba et al., 2021; Akbar, 2020; Phelps, 2024), leading to a surge in crisis response programs designed to reduce police involvement in mental health, substance use, and quality of life–related emergencies. However, these programs vary widely in structure and implementation, making it difficult to assess their impact. Some (e.g., Eugene's CAHOOTS, Denver's STAR) serve as full police alternatives, dispatching unarmed civilian teams (Dee and Pyne, 2022; Davis et al., 2024), while others (e.g., Minneapolis's Behavioral Crisis Response) operate under the co-response model, pairing mental health professionals with officers. Many of the programs differ in how they handle crises: Some (e.g., Hous-

ton, Philadelphia) resolve emergencies remotely via 911 call diversion. Others (e.g., Albuquerque Community Safety, Austin's Expanded Mobile Crisis Outreach Team) provide in-person intervention through community-based teams. Dispatch methods also vary. Most programs rely on 911 screening, but some cities, such as Rochester, New York, use a dedicated mental health crisis line (211). With over 100 programs now in operation across the US (Community Safety Workgroup, 2025), there is still limited understanding of their effectiveness, how they interact with existing emergency services, and which models best serve different communities. These gaps highlight the need for systematic evaluation of alternative crisis response strategies.

2.2 Durham's HEART

HEART Programs On June 27, 2022, the Durham Community Safety Department launched HEART to provide alternative responses to nonviolent mental health crises and quality of life–related concerns, with the aim of reducing unnecessary police involvement. HEART consists of four units: Crisis Call Diversion, Community Response Teams, Co-Response, and Care Navigation. Before the implementation of HEART, all calls were assigned to the police. Because Care Navigation focuses on follow-up services and our analysis centers on immediate 911 responses, particularly differentiating between phone and in-person immediate interventions, so this arm of HEART is not included in our analysis. The three key HEART programs in our study are described below:

1. Crisis Call Diversion, which we refer to in the following as the "remote civilian" response, embeds mental health clinicians within the 911 call center to assess behavioral health–related calls and divert nonemergent cases from requiring an in-person response. Clinicians can provide crisis deescalation, develop safety plans, and connect individuals to appropriate care without dispatching responders to the scene. If an on-scene response is needed, Crisis Call Diversion will dispatch the appropriate team, be it the Community Response Team, the Co-Response, fire/EMS, or police. Similar programs exist in Houston, Charleston, Austin, and Philadelphia.

2. Community Response Team, which we refer to as the "on-scene civilian" response, provides on-scene crisis intervention through unarmed three-person teams made up of a mental health clinician, a peer support specialist, and an EMT to handle behavioral health needs such as intoxication, trespassing, or welfare checks. By deescalating crises on-site and connecting individuals to services, The Community Response Team minimizes police contact and prevent unnecessary emergency room visits. Similar models exist in San Francisco, Denver, Portland, and Albuquerque.

3. Co-Response Team, which we refer to as "co-response," pairs a licensed clinician with a crisis intervention team (CIT)–trained police officer to respond to high-risk behavioral health crises, such as domestic violence or suicide attempts, where immediate safety concerns or weapons

may be involved. The police officer secures the scene while the clinician provides therapeutic care and follow-up services. Durham's model is distinct from the programs in many other cities, where co-response programs are housed within police or fire departments, in that it partners the Durham Country Safety Department with law enforcement.

Because one HEART team operates remotely (Crisis Call Diversion) and the two others require on-site intervention (Community Response Team, Co-Response Team), HEART's launch provides a natural setting to assess how different crisis response models affect public safety outcomes. By examining how the presence or absence of police, EMTs, and clinical specialists influences incident resolution, Durham's experience offers insights into the effectiveness of alternative crisis response strategies and their potential to shift cases away from traditional law enforcement.[3]

Background on Durham Durham's HEART program offers a unique setting in which we can evaluate multiple crisis response models within a single system. In contrast to most other cities' programs, which feature either co-response or nonpolice alternatives, HEART integrates remote civilian responses over the phone, on-scene civilian response teams, and co-response teams, where a police officer is paired with a civilian, under one framework. This structure allows direct comparisons between in-person and phone-based interventions and between police-involved and nonpolice responses, providing a more comprehensive assessment of crisis response strategies.

The spatial distribution of HEART responses is key to assessing whether the program reaches the communities most affected by crime and crisis-related incidents. Figure 2 shows that the city's Black and Hispanic residents are concentrated in central and eastern Durham, where violent crime rates are also highest.[4] HEART responses are most frequent in these areas, aligning with the greatest need for crisis intervention. The different response models are deployed distinct geographic patterns: the co-response team is densest in high-crime areas, while the remote civilian response is more evenly distributed but remains concentrated in moderate-crime areas. On-scene civilian responses are the most dispersed, reaching both high-crime neighborhoods and lower-density residential areas. The overlap between HEART responses and historically policed communities suggests the program effectively targets high-need areas. The on-scene civilian response's expanding footprint over time may indicate a growing reliance on nonpolice crisis intervention, extending the alternative response models beyond the most crime-dense neighborhoods.

[3]For more details, see DCSD.

[4]Durham's demographics further set it apart—over 45% of its residents are Black or Latino, in contrast to the demographic composition of cities with comparable programs such as Eugene's CAHOOTS (78% white) and Denver's STAR (53% white). Our study in a more racially diverse setting ensures that the evidence on crisis response effectiveness is not limited to predominantly white populations, offering insights into the equity and scalability of alternative public safety models.

3 Data and Descriptives Facts

3.1 Data Construction

Source The primary data for this analysis come from public records requests and Durham's Open Data portal, covering calls for service, HEART assignments, use-of-force incidents, and arrests. The call, HEART response, arrest, and use-of-force records extend through late 2024. Each service call record includes timestamps, call priority level, and nature codes, allowing us to identify HEART-eligible incidents. Geographic details—district, beat, census tract, and exact address—provide spatial context, while disposition codes indicate how the cases were resolved. The HEART assignment data specify which HEART program—the remote civilian response, on-scene civilian response, or on-scene co-response—was dispatched, enabling us to distinguish remote civilian responses, on-scene civilian responses, and co-responses where a police officer is paired with a civilian. The arrest data include arrestee demographics, allowing further analysis of enforcement patterns.

To complement these records, we incorporate incident data from Durham Police Department's (DPD's) National Incident-Based Reporting System (NIBRS), also accessed via Durham's Open Data portal. NIBRS tracks crimes reported to law enforcement and allows us to link incidents with calls for service by means of case numbers, providing details on whether a crime was reported and whether an arrest occurred. Each reported crime is classified as violent, property, or nonindex based on offense type. Violent crimes include robbery, homicide, murder, assault, and rape, while property crimes cover larceny, arson, burglary, and motor vehicle theft. Offenses outside these categories are grouped as nonindex crimes, which generally represent less severe offenses.

Finally, for each address in the dataset, we link the corresponding census block group data from the 2019 5-Year American Community Survey (ACS). This provides additional demographic and socioeconomic context at the census block group level, including information on race, ethnicity, unemployment rate, household income, and median rent.

Identifying Treatments and Comparison Group We identify HEART-eligible cases by combining the nature code in 911 data with HEART response records, allowing us to track which calls qualified for HEART intervention and which actually received it. This lets us compare the counterfactual outcomes—those of cases that would have been eligible for a HEART response before HEART's launch in summer 2022—with those under the actual implementation. Our treatment group includes calls classified as remote civilian response, on-scene civilian response, and/or on-scene co-response, which reflect different levels of crisis severity and need for a police or alternative response. We also identify cases where police, rather than HEART, responded to eligible calls, providing a second treatment group to assess whether HEART and police handle crises differently.

Our comparison group consists of ineligible cases, where police involvement is less certain. These include calls about suspicious activity, unknown problems, and assist-person requests, which

often require initial assessment before need for law enforcement can be determined. To ensure a relevant comparison, we exclude cases where police presence is clearly necessary, such as violent crimes. This approach provides a cleaner baseline for us to evaluate HEART's impact by focusing on calls for which alternative responses may be appropriate.

Sample Selection We begin our sample construction by processing DPD's 911 call data from January 2021 to October 2024, integrating all relevant call details, response types, and demographic information. The Calls missing essential information—such as addresses, coordinates, or disposition—are removed to ensure accuracy. The dataset is structured to track the timeline of each call, from the initial report to dispatch and resolution, with derived variables capturing response times and call duration.

To identify cases eligible for the HEART program, we apply text-based classification to the call nature codes, isolating mental health crises, welfare checks, suicide threats, and other crisis-related incidents. Calls explicitly requiring police intervention (e.g., violent crimes) are categorized separately to maintain comparability. We further verify HEART eligibility by identifying cases assigned to HEART-designated units (remote civilian response, an on-scene civilian response, and/or on-scene co-response). The final sample is restricted to cases occurring within designated police districts and beats where HEART operates. Additionally, census-based demographic indicators, including racial composition, unemployment rates, and median income, are merged to provide a richer contextual analysis.

The final dataset consists of 161,054 calls, including 43,161 ineligible cases and 103,256 HEART-eligible cases with police response and 14,637 HEART-eligible cases with HEART responses.

3.2 Descriptives

This section examines the implementation and utilization of police alternatives, tracking their evolution, impact on public safety, and variation across neighborhoods by demographics and crime rates. Using linked administrative data, we document shifts in response patterns and the extent to which these programs replace traditional policing.

Trends Over Time Figure 3 presents the evolution of call-for-service responses in Durham from 2021 to 2024, illustrating shifts in call volume and response patterns before and after the introduction of HEART in mid-2022. The top panel shows total monthly calls, distinguishing ineligible and eligible cases and responses by police and HEART teams. Before HEART's launch, police handled all eligible cases, which ranged in number from 1,800 to 2,500 calls per month, while the number of ineligible calls remained steady at 1,000 to 1,200. After HEART's introduction in mid-2022, total eligible calls increased, driven by the stability in the number of police responses at approximately 2,000 calls per month and the growing number of HEART-handled incidents. By early 2024,

HEART teams were responding to over 500 calls per month, contributing to a sustained rise in total call volume. Meanwhile, volumes of ineligible calls remained unchanged, suggesting that HEART's expansion reshaped emergency response without displacing police activity.

The bottom panel decomposes the HEART responses into nonpolice alternatives (remote civilian response and/or an on-scene civilian response) and police-involved responses, showing a clear shift toward nonpolice crisis intervention. The on-scene civilian response surpassed other response types, exceeding 500 calls per month by early 2024, largely because of the program's expansion, which increased its capacity to handle in-person crisis situations. Call volumes for remote civilian response, whereby crises are addressed remotely via phone, remained stable at lower levels, reflecting the limited scope of this type of intervention. In contrast, for co-response, whereby HEART clinicians are paired with CIT-trained officers, call volumes remained relatively low (150–200 calls per month) and fluctuated over time, suggesting police involvement is reserved for select cases. The trends indicate a growing reliance on on-scene civilian response over remote civilian response, with the on-scene civilian response team's expansion driving the overall increase in HEART responses, while the co-response's lower, stable call volume highlights its more selective use.

Summary Statistics Table 1 summarizes the ineligible cases, police-handled eligible cases, and HEART-handled eligible cases. Column (1) reports overall means, Columns (2)-–(4) break down the cases by type, Column (5) shows mean differences between HEART and police-handled cases, and Column (6) provides p-values for these differences.

Demographic and economic characteristics are broadly similar across groups, with small differences in racial composition and median household income. Eligible cases tend to occur in lower-income areas than those where ineligible cases occur, with little variation between HEART- and police-handled cases. Differences in racial composition are minor and not statistically significant.

Crime-related outcomes differ notably. Crime reports and arrests are significantly lower for HEART responses than for police-handled cases, particularly for violent crimes. Operationally, HEART cases are likelier to require backup but have shorter dispatch and response times, while clearing times remain similar. These patterns suggest that HEART focuses on noncriminal, high-need cases that require coordination but rarely result in arrests or formal crime reports.

4 Empirical Strategy

To assess HEART's impact on public safety, we compare outcomes between HEART-handled calls and police-handled calls. Table 1 demonstrates that both types of responses occur in similar neighborhoods. However, police responses to HEART-eligible cases may not serve as a valid control group due to potential behavioral adjustments. Officers might modify their enforcement practices in response to HEART's presence—either by reducing arrests in cases where HEART lacks authority

or by maintaining their previous enforcement patterns. Therefore, we use ineligible calls as our comparison group when analyzing outcomes for either police or HEART responses to eligible cases.

We follow Roth et al. (2023) and use a standard DiD as HEART rollout in June 2022, and our design is not staggered. To isolate the impact of response $r \in \{\text{Police, HEART}\}$ and each program $r \in \{\text{Civilian Remote, Civilian On-Scene, Co-Response}\}$, we estimate the following models separately for each response:

$$Y_{ict} = \alpha^r + Post_t \cdot Response_i^r \cdot \beta^r + Eligible_i^r \cdot \lambda^r + \gamma_t^r + X_{ic}' \cdot \delta^r + \epsilon_{ict}^r \tag{1}$$

where Y_{ict} represents crime reports, arrests, response time, or backup requests for call i at address c and time t. The dummy variable $Response_i^r$ equals one if $r \in \{\text{Police, HEART}\}$ or each program $r \in \{\text{Civilian Remote, Civilian On-Scene, Co-Response}\}$ responded to the call and zero otherwise. The vector X_{ic} includes covariates from Table 1, along with time-of-day, priority, and location fixed effects. Month–year fixed effects γ_t^r account for broader trends, and errors are clustered at the address level. To account for broader policing trends, we use ineligible cases—where police remained the sole responders—as a control. This ensures the observed changes reflect HEART's introduction rather than preexisting differences between eligible and ineligible areas. Here, $Eligible_i^r$ denotes HEART-eligible calls, while $Post_t$ equals one after HEART's implementation and zero otherwise.

Our key parameters, β^r, measure how the outcomes of eligible calls changed post-HEART under $r \in \{\text{Police, HEART}\}$ responses or each program $r \in \{\text{Civilian Remote, Civilian On-Scene, Co-Response}\}$, respectively. Identification again relies on the parallel trends assumption that absent $r \in \{\text{Police, HEART}\}$, the trends in police responses would have followed a similar trajectory for eligible and ineligible calls. To validate this assumption, we estimate a dynamic version of equation 1, allowing the treatment effects to vary over time:

$$Y_{ict} = \alpha^r + \sum_t D_t \cdot Response_i^r \cdot \beta_t^r + Eligible_i^r \cdot \lambda^r + \gamma_t^r + X_{ic}' \cdot \delta^r + \epsilon_{ict}^r \tag{2}$$

where D_t are month–year dummies, with May 2022 taken as the reference period. This specification compares the outcomes of eligible and ineligible cases before and after program r's launch, ensuring that preexisting differences do not drive results. If β_t^r is small and statistically insignificant for the pre-period, this would offer evidence that the parallel trends assumption holds.

Our DiD estimator for each response r measures the average treatment on the treated (ATT) the change in outcomes for HEART-eligible cases if they had been handled by police instead of response r. Specifically, the ATT quantifies the average change in outcomes that HEART-eligible cases would experience if they were handled by police instead of r. In other words, it estimates

how shifting from a system where r is available as an alternative to one where all cases default to police response affects key public safety metrics. This provides a direct measure of the impact of replacing traditional law enforcement with civilian-led crisis intervention teams.

5 Police vs. HEART Responses

5.1 Main Results

5.1.1 Impact of HEART on Incident Outcomes

Our main results refer to impact of HEART on key incident outcomes as estimated with our DiD design and presented in Figure 4. The introduction of HEART led to substantial shifts in crime reporting, arrests, response time, and need for backup, with the effects evolving over time. Importantly, the pretrends for the months leading up to HEART's implementation appear flat across all outcomes, supporting the parallel trends assumption necessary for causal interpretation.

The estimates for police-handled cases remain stable for the duration of the period, indicating that how officers respond to eligible cases did not significantly change after HEART's introduction. This suggests that the observed differences are driven by the direct effect of HEART rather than changes in policing behavior. Before HEART, 16% of eligible cases resulted in a crime report, with little difference in this outcome for cases receiving a police response before and after HEART implementation (-0.003, $SE = 0.004$). In contrast, for cases with HEART responses, crime reports declined by 9.2 pp ($SE = 0.005$), 57.5%, with the effects stabilizing after an initial adjustment period. Arrests followed a similar pattern, with a baseline rate of 5%, no significant change under police response (-0.007, $SE = 0.002$), and a 2.8 pp reduction under HEART ($SE = 0.003$). These declines were immediate and sustained, suggesting that HEART systematically handles cases in ways that reduce formal enforcement actions. Importantly, the absence of an increase in police-initiated crime reports or arrests suggests that these reductions reflect actual shifts in case handling rather than substitution effects.

Response times also differed. Before HEART, the average total time on scene was 50.01 minutes; post-HEART, this figure declined by 3.03 minutes ($SE = 0.769$) for police responses and 5.892 minutes ($SE = 1.03$) for HEART responses. The time savings were largest in HEART's early months and stabilized as the program scaled. However, calls for backup show the opposite trend. While only 3% of cases required backup before HEART, after it, this figure declined slightly for police responses (-0.004, $SE = 0.001$) but increased by 4.1 pp ($SE = 0.003$), more than doubling the baseline rate, for HEART responses. This spike was most pronounced early on and gradually attenuated, suggesting that HEART teams initially relied more on additional personnel but adjusted their approach over time.

Impact of HEART on Crime, Arrests, and Response Times Next, we break down the overall changes in crime reporting, arrests, and response times after HEART's introduction in Figure 5. Overall, the estimates for the pre-HEART trends are small and statistically nonsignificant across subcategories, supporting the parallel trends assumption.

Crime reports and arrests fell after HEART's implementation, with the largest effects for violent and other crimes. Before HEART, 8% of eligible cases involved a violent crime report; after HEART, this figure declined by 0.8 pp for police responses and by 4.6 pp for HEART responses. Arrests for violent crime, with a 3% baseline rate, fell by 0.5 pp after HEART for police responses and 1.6 pp for HEART responses. Property crime reports dropped by 1.0 pp from their preprogram baseline with HEART responses, while there was little change under police responses. Other crime reports, covering 9% of cases, declined by 5.1 pp for HEART responses, with no meaningful change for police responses. Arrest patterns followed similar trends, with HEART driving larger reductions, particularly for nonindex crimes. HEART also changed response times. Dispatch times dropped sharply, falling by 37%, while clearing times saw modest declines. The largest reductions occurred early in the rollout before stabilizing. These results suggest that HEART reduced crime reports and arrests while altering the speed and nature of emergency response.

5.1.2 Impact on Repeat 911 Calls

Next, we examine how the different crisis response models affect future 911 calls. If HEART discourages help-seeking, it could indicate mistrust in public safety authorities or unmet needs. If the public remains just as likely—or likelier—to call in their HEART-eligible cases again, this may suggest the program builds trust in emergency services rather than deterring engagement.

Figure 6 shows that in the short run (1–31 days) after either a HEART or a police response, there is no meaningful change in the likelihood of another call. Across bins of 1 day, 2-–7 days, and 8–31 days, the estimated effects are near zero and statistically insignificant, suggesting that crisis recurrence is unaffected in the immediate aftermath by the type of first responder. In the long run ($>$31 days), the likelihood of a new call is higher than its pre-program baseline for both HEART and police responses (police: 3.3 pp; HEART: 4.1 pp). Rather than signaling program inefficacy, this likely reflects continued reliance on emergency services among individuals in crisis. Crucially, HEART does not reduce future 911 engagement, indicating it does not deter people from seeking help. Instead, it may foster trust, keeping individuals connected to formal support systems without increasing punitive interactions. This distinction indicates the potential for community-oriented crisis intervention to maintain public confidence in emergency services while offering a nonenforcement approach.

5.1.3 Heterogeneity by Nature of the Call

HEART's effectiveness depends on how different emergencies are handled, as calls vary in risk, urgency, and need for enforcement vs. deescalation. Some require crisis intervention, while others may need police involvement. A heterogeneity analysis by call type can clarify whether HEART diverts cases from the criminal justice system, increases formal enforcement, or introduces tradeoffs in safety and efficiency. Figure 7 presents estimates by call type: mental health, trespass or nuisance, disturbance, violent or domestic violence incidents, and other types.[5]

HEART responses consistently reduce crime reporting, with the largest declines for the "other types" category (-5.7 pp, $SE = 0.9$) and trespass or nuisance (-3.6 pp, $SE = 0.5$). The reductions for disturbance and mental health calls are smaller, while the effects on violent or domestic violence incidents are negligible. In contrast, police responses increase crime reporting from their pre-program baseline, particularly for the "other types" category (7.3 pp, $SE = 0.9$) and violent or domestic violence incidents. Arrest patterns follow a similar trend: A HEART response significantly reduces arrests, particularly for trespass or nuisance (-2.2 pp, $SE = 0.4$) and the "other types" category (-2.4 pp, $SE = 0.5$), with smaller effects for disturbances and violent incidents. By contrast, we observe little change from baseline in arrests after police responses, which suggests that enforcement patterns remain largely unchanged.

While HEART responses reduce arrests and crime reports, they increase backup requests, particularly for disturbance (4.7 pp, $SE = 0.9$) and mental health calls (6.7 pp, $SE = 0.5$), which highlights a key operational tradeoff. HEART responses shift cases away from formal criminal processing but require greater coordination and resources to manage crisis situations effectively.

5.2 Arrestee Profiles and HEART

Figure 8 examines the impact of HEART on arrestee characteristics, comparing HEART and police responses. Arrest data provide insight into who comes into contact with law enforcement and are critical for assessing whether the alternative crisis response models shift arrest patterns. Given the well-documented disparities in policing, evaluating whether HEART changed these dynamics is key. The event-study estimates in Figure A.1 confirm no differential pretrends, supporting the validity of our DiD approach.

The top panel shows pre-HEART arrest rates by demographic group: Black individuals, men, and people aged 25–39 had the highest baseline rates, while Hispanic, White, female, and younger individuals had lower rates. The bottom panel presents DiD estimates, showing that HEART significantly reduced arrests among Black individuals (-2.4 pp, $p < 0.01$), men (-2.1 pp, $p < 0.01$), and people aged 25–39. The effects for other groups are smaller and not statistically significant, suggesting that HEART's largest impact is on populations historically overrepresented in arrests. These

[5] The "other types" category includes harassment, threats, drug-related incidents, involuntary commitment, and prostitution, spanning a wide range of severity, which makes interpretation of the estimates more complex.

declines indicate that HEART may help address policing disparities, particularly for individuals aged 25–39. The smaller effects for other groups may reflect differences in system interactions or gaps in HEART's coverage. Arrests drop more under a HEART response, suggesting that traditional policing may result in unnecessary law enforcement contact, especially for groups with higher baseline arrest rates. The demographic composition of the reference group, individuals processed by police, confirms that Black and male individuals remain most affected, reinforcing how alternative crisis response models can reduce law enforcement contact in historically heavily policed populations (Derenoncourt, 2022; Cox et al., 2024).

5.3 Robustness

5.3.1 DiD Assumptions

Bounds on the Relative Magnitudes for DiD Figure A.7 presents robustness checks on HEART's estimated impact using the bounding approach from Rambachan and Roth (2023), assessing sensitivity to deviations from parallel trends. The x-axis represents increasing restrictions on post-treatment effects, while the y-axis shows treatment estimates with 95% CI. Solid lines indicate HEART responses, and dashed lines represent police responses.

Across all outcomes—crime reports, arrests, response time, and backup requests—bounds confirm the robustness of our DiD estimates. The crime report reduction remains significant under conservative assumptions. Arrests also decline, with bounds ruling out large positive effects. Response time decreases, while backup requests rise for HEART, though without evidence of overestimation. For police-handled cases, estimates are noisier, with confidence intervals widening as restrictions increase, making their effects on crime reports and arrests less precise.

Overall, HEART's effects remain negative and meaningful across reasonable assumptions, reinforcing its role in reducing enforcement without compromising safety. Police response estimates are more sensitive to violations of parallel trends, suggesting greater uncertainty. However, the broader pattern confirms that HEART consistently lowers crime reports, arrests, and response times, while police effects are less clear.

Event Study and Hypothesized Trends Figure A.3 assesses potential violations of the parallel trends assumption following Roth (2022). The methodology accounts for pretesting concerns and low power by generating hypothesized deviations from parallel trends at 80% power. The results show no significant pre-trends in crime reports, arrests, or backup requests, confirming that observed post-treatment effects are not driven by pre-existing differences. The dashed lines represent the expected event study coefficients if deviations were present but undetectable. Notably, HEART's impact on arrests and response times remains statistically significant even under conservative assumptions about violations of parallel trends.

5.3.2 Alternative Designs

Results Table 2 presents the impact of HEART on various outcomes using multiple research designs, including DiD, triple differences, two-stage least squares (TSLS), and fuzzy DiD. To assess robustness, we compare the results from the main analysis to those from alternative research designs that recover different target parameters. While the estimates differ in magnitude, the qualitative patterns remain consistent. Column (1) reports the pre-HEART means, even-numbered columns present the treatment effects, and odd-numbered columns display the standard errors. Across all specifications, HEART responses significantly reduce crime reports and arrests. Crime reports decline by 7.9 to 9.4 pp, with the largest reductions in violent and other crimes. Similarly, arrest rates decrease, particularly for violent (-1.5 pp) and the "other crimes" category (-1.8 pp), indicating that HEART shifts cases away from formal law enforcement. These results are robust across TSLS and fuzzy DiD, reinforcing the conclusion that HEART meaningfully reduces criminal processing.

Beyond enforcement, HEART reshapes emergency response. Backup requests increase by 4.1 pp, likely reflecting the need for additional coordination in crisis intervention. Response times improve, with the total time on scene decreasing by 5 to 22 minutes, primarily because of faster dispatch and clearing times. The program also reduces the use of force by 4.8 pp, highlighting a shift toward de-escalation. While HEART does not significantly change the likelihood of a new call occurring, it does not escalate crisis recurrence. Taken together, these results suggest that HEART reduces formal legal interventions, improves response efficiency, and enhances crisis management while maintaining public safety.

6 Impact of Each Program

6.1 Remote vs. In-Person and Police-Involved Responses

To better understand how HEART affects public safety, we evaluate its impact across different response models: on-scene civilian response, remote civilian response, and co-response teams, which pairs police with mental health professionals. This distinction allows us to compare responses that fully remove police from crisis intervention with those that maintain law enforcement involvement. We employ a DiD approach, comparing the outcomes of cases with each type of response to the outcomes of ineligible cases to estimate the relative impact on crime reports, arrests, response times, and backup needs using equations 1 and 2.

Figure 9 shows that both the remote civilian response and the on-scene civilian response reduce crime reports and arrests, albeit by differing magnitudes. The remote civilian response leads to a 3.5 pp reduction in crime reports ($SE = 0.010$), while the on-scene civilian response reduces crime reports by 7.3 pp ($SE = 0.004$). Similarly, the remote civilian response and the on-scene civilian response both lower arrests, but the on-scene civilian response has a larger effect (-1.9

pp, $SE = 0.002$) compared to the remote civilian response's smaller increase (1.3 pp, $SE = 0.003$). The increase in arrests for the remote civilian cases may be linked to the initial challenges in fully resolving incidents remotely, as reflected in the surge in backup requests for remote civilian response (7.6 pp, $SE = 0.018$) in the early months. Many remote civilian responses required additional intervention by on-scene teams, leading to police involvement and potential arrests. Over time, backup requests for remote civilian responses declined, suggesting improved triage and growing confidence in remote handling of crises, which may have mitigated the need for escalation.

Response time patterns further differentiate these models. The remote civilian response significantly reduces the total response time (-25.89 minutes, $SE = 2.44$), while the on-scene civilian response increases the response time by 5.67 minutes ($SE = 1.02$) and the co-response extends response duration by 4.70 minutes ($SE = 1.55$). These findings reflect the efficiency of remote interventions and the more challenging, time-intensive nature of in-person crisis teams. However, the early increase in the remote civilian response backup requests, coupled with the smaller reduction in crime reports and increase in arrests for this response type, suggests that remote interventions faced challenges in fully resolving certain crises without additional support. In contrast, the greater reduction in crime reports and arrests under the on-scene civilian response, despite the longer response times, suggests that in-person teams were more effective in deescalating incidents on-site without law enforcement intervention.

Overall, HEART's crime prevention and arrest reduction effects are concentrated in nonpolice responses, with on-scene civilian response yielding the largest declines. The stabilization of the remote civilian response's backup requests over time suggests that initial challenges in remote crisis management improved, but the smaller impact on arrests under this response type indicates the limits of phone-based interventions in fully deescalating certain situations. Meanwhile, COR's minimal effect on crime reports and arrests suggests that when police are present, response patterns largely mirror those under traditional enforcement. These results highlight the key distinction between remote and in-person crisis response: While both reduce crime reports, in-person teams more effectively prevent escalation without requiring additional support from law enforcement.

6.2 Does HEART Lead to Escalation?

Crisis calls, if mishandled, can generate significant externalities, from community distrust to legal risks (Bor et al., 2018; Ang et al., 2024). Police responses, while necessary in some cases, carry the potential for escalation, as law enforcement holds a monopoly on state-sanctioned force. Alternatives that reduce police involvement may mitigate this risk. At the same time, an inadequate response can leave underlying issues unresolved, increasing the chance of future violent crime. Hence, Figure 10 evaluates whether HEART responses lead to lower use-of-force incidents and fewer follow-up emergency calls involving violent crime.

The left panel of Figure 10 presents DiD estimates on police use of force ($\times 100$). HEART

responses are associated with fewer use-of-force incidents than police-handled cases, with the on-scene civilian response showing the largest reduction, reinforcing our conclusion about the effectiveness of on-scene civilian response. In contrast, the remote civilian response shows weaker and more variable effects, suggesting that remote interventions may not consistently deescalate crises and may sometimes require police backup. For the co-response teams, these outcomes exhibit little change, indicating that when police are present, prior use-of-force patterns persist. These results highlight the advantages of direct, unarmed crisis response in reducing coercive encounters.

The right panel of Figure 10 shows DiD estimates on the likelihood of a future 911 call involving violent crime, with a pre-HEART mean of 4%. The on-scene civilian response drives most of the observed decline, suggesting that on-scene crisis teams effectively prevent escalation. For the remote civilian response, the estimates are also negative, but noisier, likely reflecting instances where remote responses fail to fully resolve crises, leading to police involvement. Co-response shows minimal impact: This leads us to conclude that police presence does not significantly alter the risk of escalation. These patterns suggest that in-person crisis response is more reliable in preventing repeated violent incidents while remote interventions may have limitations in fully stabilizing situations.

6.3 Does HEART Improve Trust in 911?

HEART's introduction offers a unique opportunity to examine whether expanding crisis response options changes how communities engage with 911. A decline in call priority may signal increased trust in emergency services, whereby people feel comfortable seeking help for lower-stakes situations rather than only emergencies. Research often infers trust in police from high-profile violence cases (Desmond et al., 2016; Rivera and Ba, 2023; Ang et al., 2024) or self-reported surveys (Tyler, 2001; Pickett et al., 2022; Cheng and Liu, 2025). Our setting allows us to test whether the availability of police alternatives leads to greater reliance on 911.

Figure 11 estimates the impact of the remote civilian response, the on-scene civilian response, and co-response on call priority, where higher values indicate more urgent cases. The remote civilian response and the on-scene civilian response significantly lower call priority, with the remote civilian response showing the largest decline, suggesting that remote and unarmed crisis teams expand access to emergency services for noncritical incidents. Co-response and police-only responses also see declines in call priority, albeit of smaller magnitudes, indicating that while HEART shifts the composition of calls, police-led interventions are still associated with more urgent cases but may benefit from spillover effects of increased public trust. The effects are meaningful—the remote civilian response reduces priority by more than one full point, while the on-scene civilian response's effect is smaller but still notable. These findings suggest that HEART, particularly through the remote civilian response and the on-scene civilian response, helps normalize 911 use for nonemergencies if a nonpolice response exists, reinforcing trust in crisis response services.

6.4 Labor Costs

The expansion of HEART reshaped 911 call patterns and has significant cost implications. As trust in emergency services grows, cities must weigh the financial trade-offs between alternative response models and the savings from reduced police reliance. Figure 12 highlights these differences, breaking down the hourly costs of each type of response, the scale of calls handled, and the overall labor cost impact. The top-left panel shows that police responses cost $67.53 per hour while the cost of the HEART models varies significantly: the remote civilian response is the least expensive at $32.50, the on-scene civilian response is the most expensive at $98.09, and the co-response in between at $73.90. The top-right panel reveals the distribution of calls—police responses remain dominant, with police handling 103,256 calls, compared to 8,909 for the on-scene civilian response, 2,857 for the remote civilian response, and 2,871 for co-response.

The bottom panel quantifies HEART's impact on labor costs using the DiD estimates, where we calculate the outcome by multiplying the average hourly wage for each type of response by the time required to handle the calls. We find the remote civilian response reduces costs by $48.90 per call while the on-scene civilian response and the co-response increase expenditures by $30.91 and $11.62, respectively. Even the modest reduction of $3.41 per call for police-only responses would translate to approximately $352,103 in savings across all police-handled calls on the police budget. Figure A.6 presents the event-study versions of these estimates, confirming that the pretrends before HEART implementation were similar.

The cost dynamics shift when we look at total expenditures. The remote civilian response generates the largest per-call savings because of the lower wages and smaller call volume, but its overall budget impact remains limited. The on-scene civilian response, by contrast, is the most expensive model, reflecting the city's decision to prioritize higher-touch, in-person interventions, and its expansion significantly increases labor costs. Co-response falls in between, reducing costs modestly while maintaining police involvement. The city's continued investment in on-scene civilian teams suggests a shift toward civilian-led crisis response, despite its higher price tag. While expanding the on-scene civilian response raises immediate costs, the potential savings from fewer arrests, lower incarceration rates, and reduced crime could outweigh these expenses. If HEART prevents even a fraction of arrests or violent escalations, the prevention of legal and social costs—to both individuals and the criminal justice system—may more than justify the investment. Cities face a fundamental trade-off: maximizing cost efficiency with remote interventions such as the remote civilian response or investing in in-person responses such as the on-scene civilian response that could deliver greater long-term public safety and social benefits.

6.5 Robustness

7 Welfare Implications

This section examines the public's valuation of HEART and its cost-effectiveness as a crisis response model. We estimate residents' WTP, reflecting perceived benefits such as reduced harm, better long-term outcomes, and greater public trust. To determine if these benefits outweigh costs, we compare WTP to the net fiscal burden, including direct expenditures and fiscal externalities like changes in arrests and police resource use. Using the MVPF framework (Hendren and Sprung-Keyser, 2020), we quantify HEART's economic return by measuring net social benefits per dollar of government spending.

7.1 Government Costs

To benchmark HEART's value to residents, we first establish the program's fiscal cost operating HEART, including personnel expenses, training, and administrative costs. Over the analyzed period from 2021 to 2024, the total direct cost of the HEART program amounted to $17,433,076 suggesting a $1,191 for each of the 14,637 HEART responses in our sample. This budget reflects a steady increase in funding, rising from $1.86 million in 2021 to $6.54 million in 2024 (City of Durham, 2023). Given Durham's 2024 population of 318,353 residents, the per-resident cost rose from $5.85 in 2021 to $20.54 in 2024, with intermediate costs of $15.35 in 2022 and $13.01 in 2023. These figures provide a benchmark for evaluating the program's value to residents, motivating our estimation of WTP and its comparison to HEART's net fiscal cost using the MVPF framework.

7.2 Measuring Public WTP Using Contingent Valuation Survey

Despite the rapid expansion of nonpolice crisis response programs, no existing research quantifies the public's WTP for these services, leaving a critical gap in understanding their perceived value and fiscal sustainability. To address this, we assess HEART's benefits as valued by Durham residents through their WTP, which represents the income reduction that leaves them indifferent to its presence and reflects the city's valuation of an alternative crisis response system. The WTP captures four key dimensions: (i) reduced harm to individuals in crisis, including fewer arrests and less psychological trauma; (ii) improved long-term outcomes through better mental health treatment and reduced criminal records; (iii) lower burdens on jails and courts; and (iv) increased public trust, reduced fear of law enforcement, and enhanced community well-being.

To estimate WTP, we employ a contingent valuation (CV) survey, a widely used method for valuing nonmarket goods in policy research. Unlike traditional crime control valuations, which exclude offender utility from social welfare calculations (Trumbull, 1990; Cohen, 2000; Chalfin, 2015), our approach accounts for the broader societal benefits of alternative crisis response programs. CV en-

ables us to assess benefits without relying on exogenous variation, provides an ex-ante estimate of both tangible and intangible benefits, and ensures respondents are well-informed about HEART before stating their valuations (Cohen, 2000; Esberg and Mummolo, 2018). While CV is subject to response biases, budget neglect, and concerns about sample representativeness, we adhere to best practices to minimize these limitations (Bishop et al., 2017; Domínguez and Scartascini, 2024).

Sample and Logistics We address each of these potential biases by following the CV best practices used by the federal government and other researchers (Bishop et al., 2017; Domínguez and Scartascini, 2024) along with best practices for experiments generally (Stantcheva, 2023). We recruit a sample of 143 subjects who are nationally representative on gender, age, race, and ethnicity. We recruit subjects from CloudResearch instead of using in-person surveys to reduce concerns about social desirability bias (Leggett et al., 2003). Following the guidance from Kuziemko et al. (2015), we undertake the following steps to ensure the sample's validity. First, we only allow participants with a US IP address who launched our survey on a workday morning. Second, we include a CAPTCHA to exclude potential robots. Third, we tell participants that payment is contingent on survey completion. Finally, we include attention and comprehension checks.

Survey Structure The survey began with an attention check to ensure engagement, followed by a five-digit ZIP code entry to assess geographic representation. To enhance response consequentiality (Bishop et al., 2017), participants were informed that results would be shared with local policymakers as input for public safety policies. Respondents then received background information on HEART's purpose and scope before completing a comprehension check. Incorrect responses prompted a review and retest before proceeding.

Next, the willingness-to-pay (WTP) section assessed valuation using a multiple price list format. Participants faced four binary choices, each asking if they would support HEART with an annual tax, randomly assigned between $3 and $108. This approach mitigated anchoring effects and allowed for a granular WTP estimate. To reduce hypothetical bias, respondents were reminded that agreeing to a tax would reduce disposable income, limiting other spending. Policymaker awareness of survey results reinforced the real-world relevance of their choices.

Analyzing WTP responses provides insights into public support for crisis response alternatives. Additional survey questions gathered demographic information and views on policing and crisis response, including preferences for police alternatives in scenarios from Ba et al. (2024), contextualizing support for HEART within broader public safety attitudes.

Computing the WTP To compute the WTP, we proceed to use subject's answers to the four pairwise choices where she must report whether she would vote for the HEART program. These answers map to WTP bins. The WTP for HEART can fall into one of ten intervals: $(-\infty, 3]$, $(3, 12]$, $(12, 21]$, $(21, 34]$, $(34, 46]$, $(46, 56]$, $(56, 69]$, $(69, 81]$, $(81, 94]$ $(94, 108]$, $(108, \infty)$. We use nonfocal numbers

to prevent subjects from bunching at round numbers (Bishop et al., 2017). Figure 13 presents the distribution of WTP. We construct the WTP from the multiple price list responses using the same methods as Allcott and Kessler (2019) and Holz et al. (2024). We assign one unique WTP for each range. For the eight interior ranges, we assign the mean of the endpoints. For example, we assign a WTP of \$7.5 for all responses on $(3,12]$ and a WTP of \$75 for all responses on (69,81]. For the unbounded ranges, we assume that the conditional distribution of the WTP is triangular.[6] This gives \$-10.50 as the conditional mean on $(-\infty,3)$ and \$127.11 as the conditional mean on $(108,\infty)$.

WTP Estimates Using this method, we estimate a mean WTP for HEART of \$67.50 per respondent per year or WTP \$5872 for each HEART response. When restricting the sample to the 134 of 143 respondents who reported understanding the survey questions, the mean WTP increases slightly to \$69.77. Notably, about 86% of participants express a positive WTP, and a substantial majority, 73%, report a WTP exceeding the highest per-resident cost of \$20.54 in 2024. Even at the lowest per-resident cost of \$5.85 in 2021, WTP remains well above the program's direct cost across all years, including intermediate costs of \$15.35 in 2022 and \$13.01 in 2023. Furthermore, 78% of respondents have a WTP greater than HEART's net fiscal cost. These results indicate strong public support for police alternatives, with WTP consistently exceeding program costs, even before incorporating fiscal externalities.

7.3 Measuring Cost-Effectiveness with MVPF

Overview of the MVPF The MVPF quantifies a policy's cost-effectiveness by comparing the benefits individuals receive to the net fiscal cost. It captures how much net social benefit HEART generates per additional dollar of government spending, offering a rigorous framework to evaluate its overall welfare impact. The MVPF is defined as:

$$MVPF = \frac{\text{Beneficiaries' Willingness to Pay}}{\text{Direct Cost of HEART} - \text{Fiscal Externality}} \tag{3}$$

The numerator represents the WTP for HEART, which accounts for both individuals directly served by HEART and broader community members who benefit from the program's availability. This value should be weakly positive, unless the public perceives HEART as net harmful—a scenario inconsistent with the survey evidence we present later.

The denominator consists of two components. The first is the direct cost of operating HEART, including personnel expenses, training, and administrative costs. The second is the fiscal externality, which captures broader financial spillovers, such as savings from fewer arrests, reduced incarceration costs, and changes in police resource allocation. These externalities can be negative (leading

[6] We assume that the initial density is equal to the average density on the adjacent range and then solve for the upper bound.

to government cost savings) or positive (creating additional expenses for the public sector). The net cost to the government is the sum of these two components.

The MVPF provides an intuitive benchmark for policy evaluation. If the MVPF exceeds one, it indicates that HEART's benefits exceed its costs, making it a net positive investment. If a positive fiscal externality exceeds the direct cost of the program and the WTP is positive, then we say that the $MVPF = \infty$ as the policy fully funds itself through fiscal feedback. In other words, it suggests that fiscal savings fully offset HEART's direct costs, meaning that the program effectively funds itself through reductions in criminal justice expenditures.

Fiscal Externality Motivated by the consistent finding that HEART reduces criminalization and arrests—leading to cost savings from both averted crimes and reductions in public safety expenditures—we incorporate these effects by constructing an index that captures the fiscal externalities associated with changes in public safety costs.

To calculate the fiscal externality of HEART, we account for the government's costs and savings, using estimates from Miller et al. (2021) on the economic costs of crime in the U.S. for 2017, including tangible (e.g., medical, productivity, property, public services) and intangible (quality of life) costs. We report those estimates in table A.2 of the Appendix. For each call i with response $r \in$ Police, Civilian Remote, Civilian On-Scene, Co-Response and crime type k, we compute total costs as follows:[7]

$$\begin{aligned} \text{Total Costs}_i^{r,k} = \text{Crime}_i^k \times (&\text{Medical}_i^k + \text{Mental Health}_i^k + \text{Productivity Loss}_i^k \\ &+ \text{Property Loss}_i^k + \text{Public Services}_i^k + \\ &+ \text{Adjudication \& Sanctioning}_i^k + \text{Perpetrator Work Loss}_i^k) \end{aligned} \tag{4}$$

where Crime_i^k represent whether a crime of type k was reported. The fiscal externality of HEART is captured by the change in Total Costs following its introduction.

To estimate the average change in total costs following the introduction of HEART, we employ the DiD model from equation 1, where the dependent variable is $\text{Total Costs}_i^{r,k}$. This approach recovers $\hat{\Delta}_{HEART}^{\text{Costs}}$, capturing the program's impact on cost savings or increase from HEART responses on eligible cases, i.e., the fiscal externality.

MVPF Estimates . Figure 14 illustrates the impact of HEART's introduction on cost per call , capturing both direct program costs and broader fiscal externalities. Our analysis reveals an overall cost reduction , with the largest savings driven by quality of life improvements, reductions in tangible costs, and lower perpetrator work loss , leading to total crime cost savings of approximately $2,092 . In contrast, the direct operational cost of a HEART response is $1,191 , which is more

[7] For clarity, we omit subscripts for address c and time t.

than offset by these fiscal savings, resulting in a net government cost impact of -$901 . This suggests that HEART delivers substantial fiscal benefits by reducing crime-related expenses, including victimization, investigation, and incarceration costs. Furthermore, with a per-response willingness to pay (WTP) of $5,872 , the program generates positive net benefits , implying an MVPF of infinity (∞) , as fiscal savings fully cover its operational costs.

8 Conclusion

This paper evaluates Durham's HEART program, providing evidence that civilian-led crisis response reduces crime reports, arrests, and response times— primarily driven by fully civilian teams — while police-clinician co-response has minimal impact. Alternative response models shift nonviolent incidents away from police without increasing crime, and concerns over backup requests diminish over time. Using a novel contingent valuation survey, we find strong public support, with willingness to pay exceeding HEART's net cost. Applying the Marginal Value of Public Funds framework, we find that HEART's fiscal savings exceed its direct costs, indicating that the program effectively funds itself through externalities.

These findings contribute to a growing policy discussion on the role of police and the need for non-law enforcement alternatives in public safety. However, key questions remain. First, assessing whether shifts in emergency response translate into broader crime deterrence or improved social outcomes remains an open question. Second, generalizability remains a challenge: Durham's centralized 911 integration allows a clean comparison of response models, but many cities operate with fragmented dispatch systems. Future research should explore whether similar gains hold in cities with different crisis response structures, as well as how alternative programs affect disparities in policing outcomes at scale.

Looking ahead, our study provides a replicable framework for evaluating police alternatives, combining administrative data, natural experiments, and public valuation methods to assess both effectiveness and fiscal sustainability. As cities continue to invest in non-police emergency response models, rigorous evaluation is critical to ensuring these programs deliver on their promise—providing both safety and a more equitable approach to crisis intervention.

References

Akbar, A. A. (2020). An abolitionist horizon for (police) reform. *Calif. L. Rev. 108*, 1781.

Allcott, H. and J. B. Kessler (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics 11*(1), 236–276.

Anderson, S., G. W. Harrison, M. I. Lau, and R. E. Elisabet (2007). Valuation using multiple price list formats. *Applied Economics 39*(6), 675–682.

Ang, D., P. Bencsik, J. Bruhn, and E. Derenoncourt (2024). Community engagement with law enforcement after high-profile acts of police violence. *American Economic Review: Insights,Forthcoming*.

Ba, B., M. Baskar, T. Cheng, and R. Mariman (2024). Understanding demand for police alternatives. Technical report.

Ba, B. A., D. Knox, J. Mummolo, and R. Rivera (2021). The role of officer race and gender in police-civilian interactions in chicago. *Science 371*(6530), 696–702.

Ba, B. A., R. R. Rivera, and A. Whitefield (2023a). Forecasting the impact of racial uprisings, market versus stakeholders' expectations.

Ba, B. A., R. R. Rivera, and A. Whitefield (2023b). Market response to racial uprisings.

Bishop, R. C., K. J. Boyle, R. T. Carson, D. Chapman, W. M. Hanemann, B. Kanninen, R. J. Kopp, J. A. Krosnick, J. List, N. Meade, et al. (2017). Putting a value on injuries to natural assets: The bp oil spill. *Science 356*(6335), 253–254.

Bor, J., A. S. Venkataramani, D. R. Williams, and A. C. Tsai (2018). Spillover effects of police killings on the mental health of black americans in the general us population. *The Lancet 392*(10144), 302–310.

Bursztyn, L., G. Egorov, I. Haaland, A. Rao, and C. Roth (2023, 01). Justifying Dissent*. *The Quarterly Journal of Economics 138*(3), 1403–1451.

Chalfin, A. (2015). Economic Costs of Crime. *The Encyclopedia of Crime and Punishment*, 1–12.

Chalfin, A. and J. McCrary (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature 55*(1), 5–48.

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

Chen, M. K., K. L. Christensen, E. John, E. Owens, and Y. Zhuo (2023, 09). Smartphone data reveal neighborhood-level racial disparities in police presence. *The Review of Economics and Statistics*, 1–29.

Cheng, T. and S. Liu (2025). How the state is seen: Public perceptions of legibility and legitimacy in chicago policing. *Working Paper*.

Chyn, E., K. Haggag, and B. A. Stuart (2022). The effects of racial segregation on intergenerational mobility: Evidence from historical railroad placement. Technical report.

City of Durham (2023). Fiscal year 2023-24 adopted budget. Pages 238-239.

Cohen, M. A. (2000). Measuring the costs and benefits of crime and justice. *Criminal justice 4*(1), 263–315.

Community Safety Workgroup (2025). *Alternative Crisis Response Programs: First of its Kind Directory of Direct-Dispatch, Non-Police Programs*. Community Safety Workgroup.

Cox, R., J. P. Cunningham, and A. Ortega (2024, August). The impact of affirmative action litigation on police killings of civilians. Working Paper.

Cox, R., J. P. Cunningham, A. Ortega, and K. Whaley (2025). Black lives: The high cost of segregation. *American Economic Journal: Economic Policy,Forthcoming*.

Cullen, Z., W. Dobbie, and M. Hoffman (2022, 08). Increasing the demand for workers with a criminal record*. *The Quarterly Journal of Economics 138*(1), 103–150.

Davis, J., S. Norris, J. Schmitt, Y. Shem-Tov, and C. Strickland (2024). Can mental health crisis-response teams support better policing?

Dee, T. S. and J. Pyne (2022). A community response approach to mental health and substance abuse crises reduced crime. *Science advances 8*(23), eabm2106.

Derenoncourt, E. (2022, February). Can you move to opportunity? evidence from the great migration. *American Economic Review 112*(2), 369–408.

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

Desmond, M., A. V. Papachristos, and D. S. Kirk (2016). Police violence and citizen crime reporting in the black community. *American Sociological Review 81*(5), 857–876.

Deza, M., T. Lu, J. C. Maclean, and A. Ortega (2023). Treatment for mental health and substance use: Spillovers to police safety. Working Paper 31391, National Bureau of Economic Research.

Domínguez, P. and C. Scartascini (2024). Willingness to pay for crime reduction: The role of information in the americas. *Journal of Public Economics 239*, 105205.

Dube, O., S. J. MacArthur, and A. K. Shah (2024, 11). A cognitive view of policing*. *The Quarterly Journal of Economics 140*(1), 745–791.

Esberg, J. and J. Mummolo (2018). Explaining misperceptions of crime. *Available at SSRN 3208303*.

Evans, W. N. and E. G. Owens (2007). Cops and crime. *Journal of Public Economics 91*(1), 181–201.

Foureaux Koppensteiner, M., J. Matheson, and R. Plugor (2024, February). The impact of improving access to support services for victims of domestic violence on demand for services and victim outcomes. *American Economic Journal: Economic Policy 16*(1), 292–324.

Gates, S., B. Ariel, and N. Assaraf (2024). Responding to nonemergency calls for service via video: A randomized controlled trial. *Criminology & Public Policy*.

Hart, O., A. Shleifer, and R. W. Vishny (1997, 11). The Proper Scope of Government: Theory and an Application to Prisons*. *112*(4), 1127–1161.

Hendren, N. (2016). The policy elasticity. *Tax Policy and the Economy 30*.

Hendren, N. and B. Sprung-Keyser (2020, 03). A unified welfare analysis of government policies*. *The Quarterly Journal of Economics 135*(3), 1209–1318.

Hoekstra, M. and C. Sloan (2022, March). Does race matter for police use of force? evidence from 911 calls. *American Economic Review 112*(3), 827–60.

Holz, J., R. Jiménez-Durán, and E. Laguna-Müggenburg (2024). Estimating the distaste for price gouging with incentivized consumer reports. *American Economic Journal: Applied Economics 16*(1), 33–59.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Kaba, M., N. Murakawa, and T. K. Nopper (2021). *We Do This 'Til We Free Us: Abolitionist Organizing and Transforming Justice*. Chicago, IL: Haymarket Books.

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review 105*(4), 1478–1508.

Leggett, C. G., N. S. Kleckner, K. J. Boyle, J. W. Dufield, and R. C. Mitchell (2003). Social desirability bias in contingent valuation surveys administered through in-person interviews. *Land economics 79*(4), 561–575.

Mello, S. (2019). More cops, less crime. *Journal of Public Economics 172*, 174–200.

Metcalfe, R. D. and S. Roth (2025). Making the invisible visible: The impact of revealing indoor air pollution on behavior and welfare. Technical report, National Bureau of Economic Research.

Mikdash, M. and R. Zaiour (2024). The impact of police shootings on gun violence and civilian cooperation. *Journal of Public Economics 237*, 105189.

Miller, T. R., M. A. Cohen, D. I. Swedler, B. Ali, and D. V. Hendrie (2021). Incidence and costs of personal and property crimes in the usa, 2017. *Journal of Benefit-Cost Analysis 12*(1), 24–54.

Moreno-Medina, J., A. Ouss P. Bayer, and B. A. Ba (2025, 01). Officer-involved: The media language of police killings*. *The Quarterly Journal of Economics*, qjaf004.

Mukherjee, A. (2021, May). Impacts of private prison contracting on inmate time served and recidivism. *American Economic Journal: Economic Policy 13*(2), 408–38.

Mukherjee, A. and N. J. Sanders (2021, July). The causal effect of heat on violence: Social implications of unmitigated heat among the incarcerated. Working Paper 28987, National Bureau of Economic Research.

Ouss, A. and M. Stevenson (2023, July). Does cash bail deter misconduct? *American Economic Journal: Applied Economics 15*(3), 150–82.

Owens, E., D. Weisburd, K. L. Amendola, and G. P. Alpert (2018). Can you build a better cop? *Criminology & Public Policy 17*(1), 41–87.

Phelps, M. S. (2024). *The Minneapolis Reckoning: Race, Violence, and the Politics of Policing in America*. Princeton University Press.

Pickett, J. T., A. Graham, and F. T. Cullen (2022). The american racial divide in fear of the police. *Criminology 60*(2), 291–320.

Rambachan, A. and J. Roth (2023, 02). A More Credible Approach to Parallel Trends. *The Review of Economic Studies 90*(5), 2555–2591.

Rivera, R. G. and B. A. Ba (2023, 10). The effect of police oversight on crime and misconduct allegations: Evidence from chicago. *The Review of Economics and Statistics*, 1–45.

Ross, M. and C. Sloan (2024). The effect of field training officers on police use of force.

Roth, J. (2022, September). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights 4*(3), 305–22.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics 235*(2), 2218–2244.

Roth, J. and P. H. C. Sant'Anna (2023). Efficient estimation for staggered rollout designs. *Journal of Political Economy Microeconomics 1*(4), 669–709.

Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics 15*(1), 205–234.

Stevenson, M. T. and S. G. Mayson (2022). Pretrial detention and the value of liberty. *Virginia Law Review 108*(3), pp. 709–782.

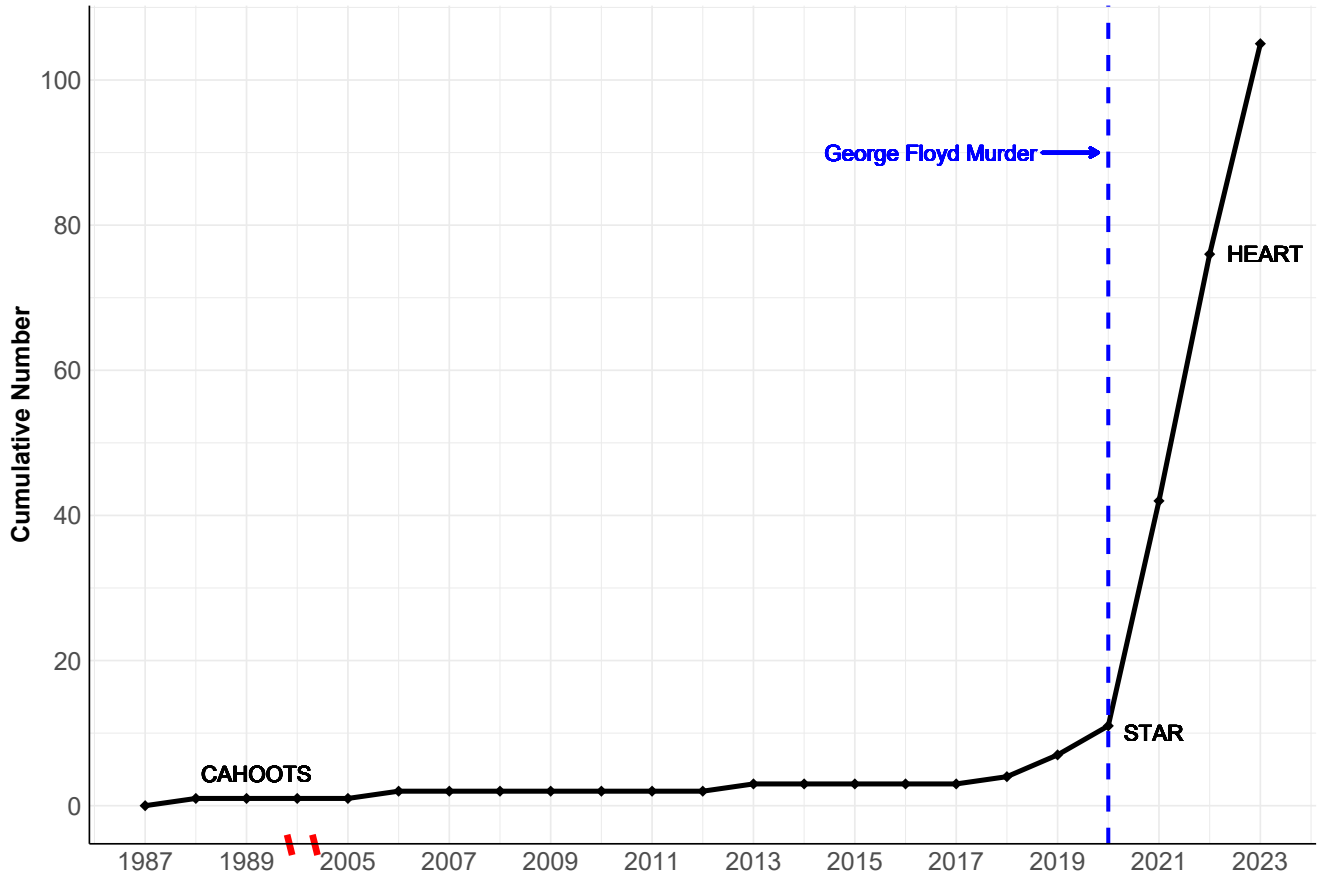
Trumbull, W. N. (1990). Who has standing in cost-benefit analysis? *Journal of Policy Analysis and Management 9*(2), 201–218.

Tyler, T. R. (2001). Public trust and confidence in legal authorities: What do majority and minority group members want from the law and legal institutions? *Behavioral sciences & the law 19*(2), 215–235.

Vidal, J. B. I. and T. Kirchmaier (2018). The effect of police response time on crime clearance rates. *The Review of Economic Studies 85*(2 (303)), 855–891.

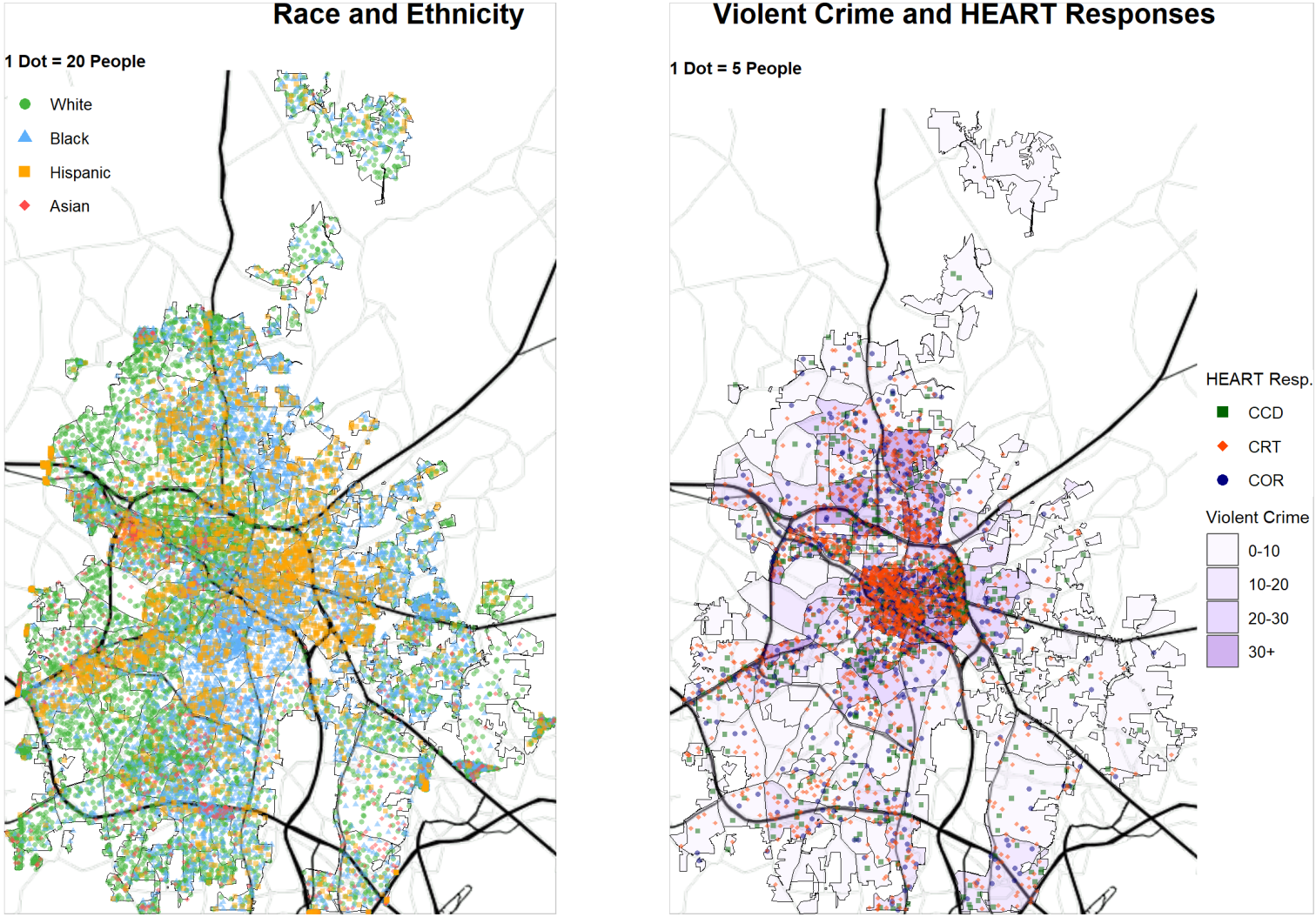
Yegen, E. (2021). Do institutional investors mitigate social costs of privatization? evidence from prisons. Technical report.

Figure 1: US Alternative Crisis Response Programs Over Time



Notes: This figure shows the growth in the number of local governments implementing behavioral health crisis response programs over time. The data are sourced from research compiled in the "Community Safety Workshop in Alternative Crisis Response Programs: First of Its Kind Directory of Direct-Dispatch, Non-Police Programs," which documents municipalities and counties that have established dedicated emergency response teams. The directory provides detailed descriptions of these programs, their structures, and their operational models. For reference, we highlight HEART in Durham, North Carolina, CAHOOTS in Eugene, Oregon, and STAR in Denver, Colorado—three programs frequently cited in the literature as key models of alternative crisis response.

Figure 2: Demographics, Crime, and HEART Responses in Durham



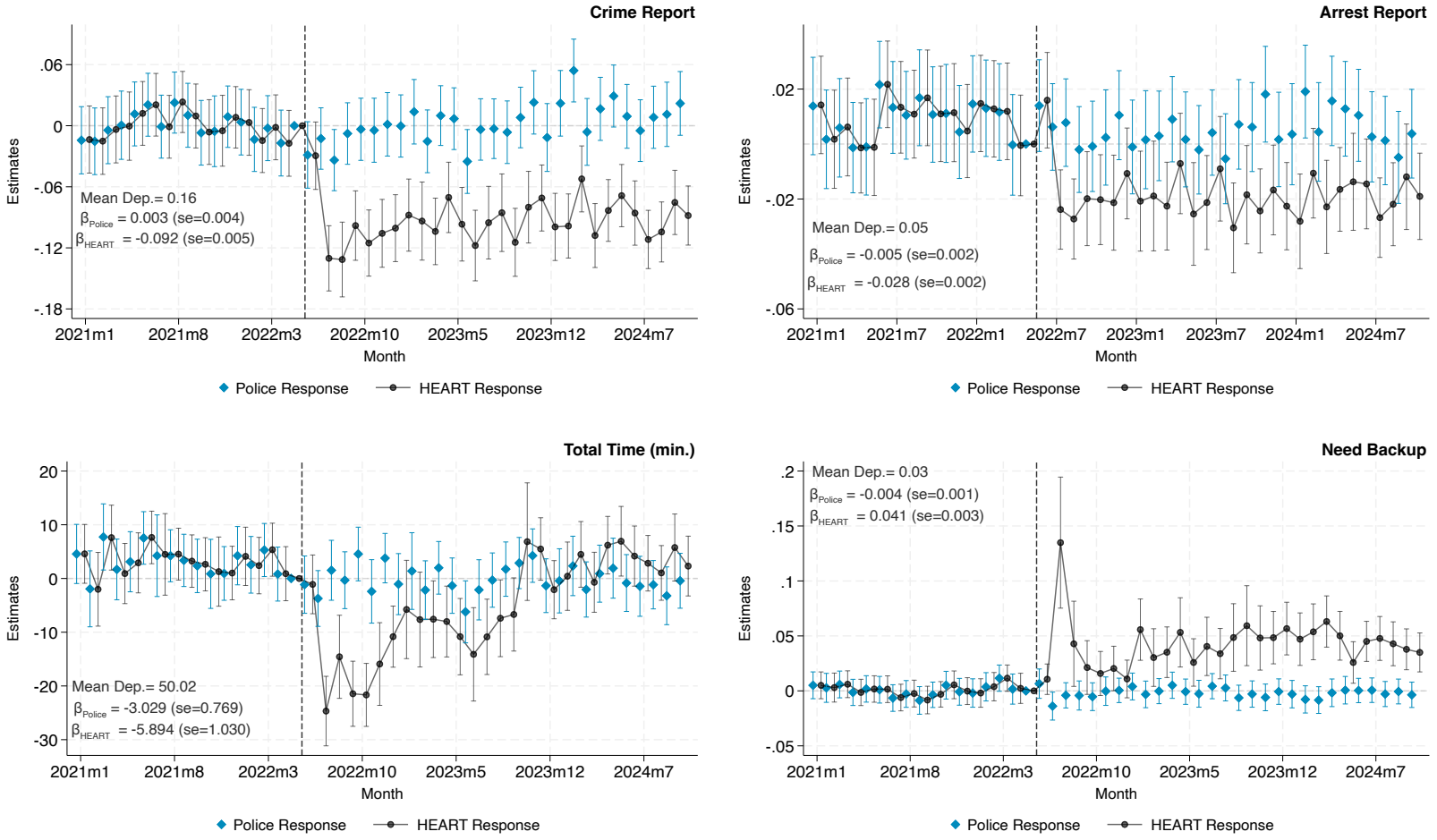
Notes: The left panel displays the racial and ethnic composition of Durham's population using 2019 ACS data at the census-block-group level, with each dot representing 20 residents. The right panel presents violent crime rates per 1,000 residents in 2019 alongside HEART response locations.

Figure 3: HEART-Ineligible Cases, HEART-Eligible Cases, HEART Responses, and Police Responses



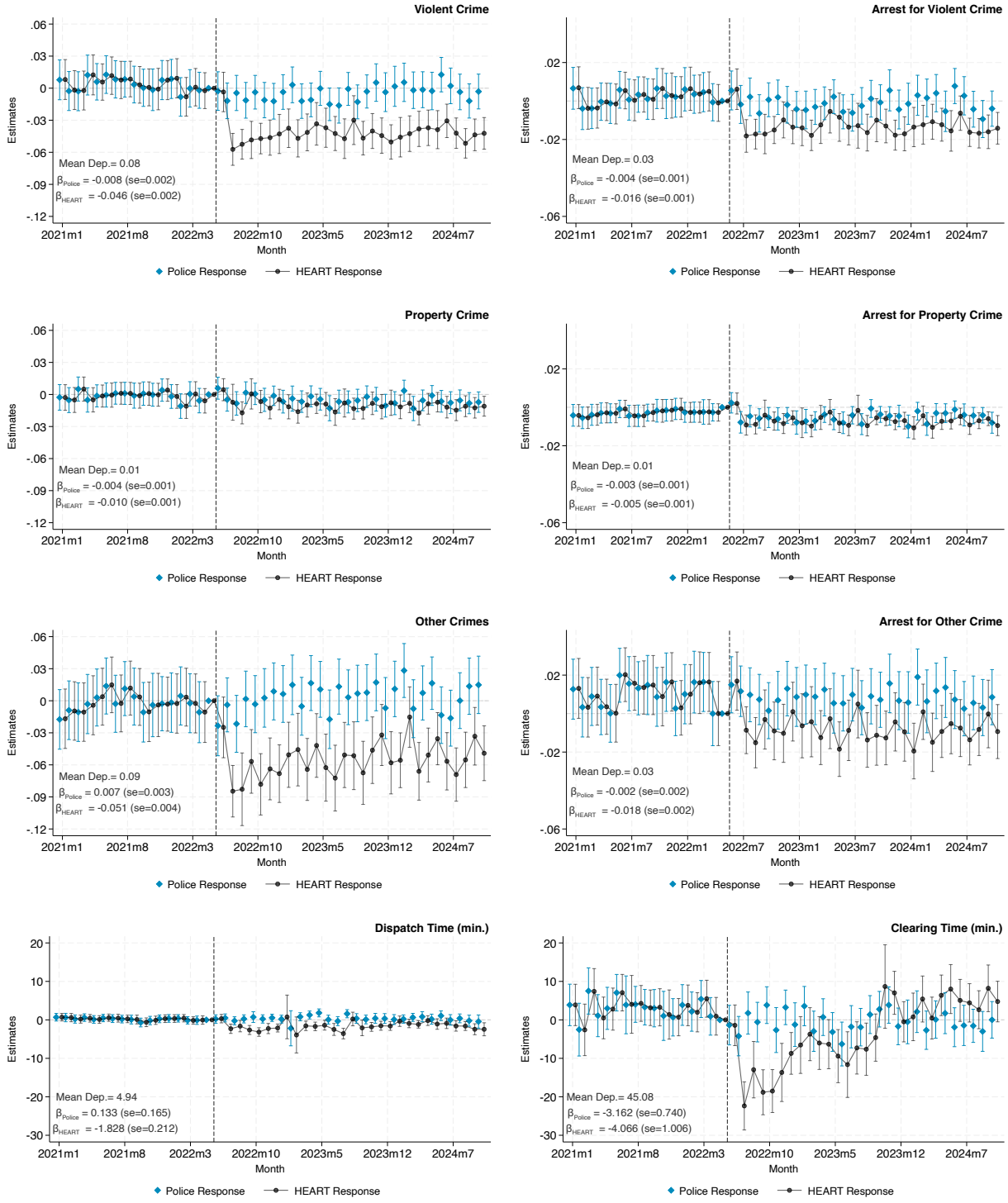
Notes: The top-left figure shows monthly call trends from 2021 to 2024, distinguishing HEART-ineligible and HEART-eligible cases and police and HEART responses. The eligible cases fall under civilian remote, civilian on-scene, and co-response. The top right panel tracks the implementation and expansion of HEART responses over time. The vertical dashed line marks the mid-2022 launch of HEART. The bottom panel indicates the nature codes for ineligible cases and each response types.

Figure 4: Impact of HEART Program on Incident Outcomes



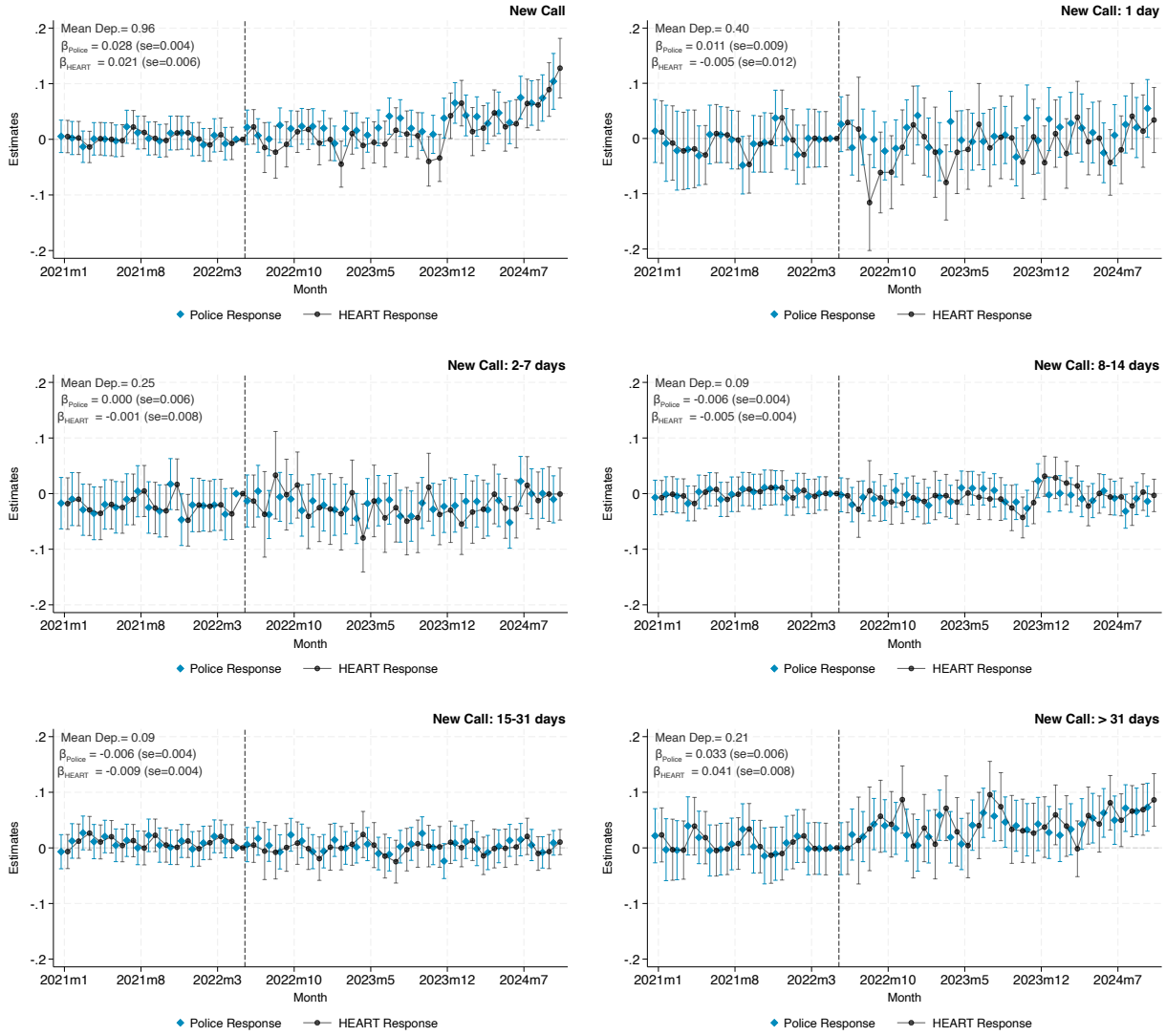
Notes: These figures present the impact of the HEART program on incident outcomes, including crime, arrests, response time, and need for backup. They present DiD estimates for police and HEART responses with coefficients and 95% confidence intervals. The mean of the dependent variable for eligible cases in the pre-period is included for reference. Regressions include day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure 5: Impact of HEART Program on Crime, Arrests, and Response Times



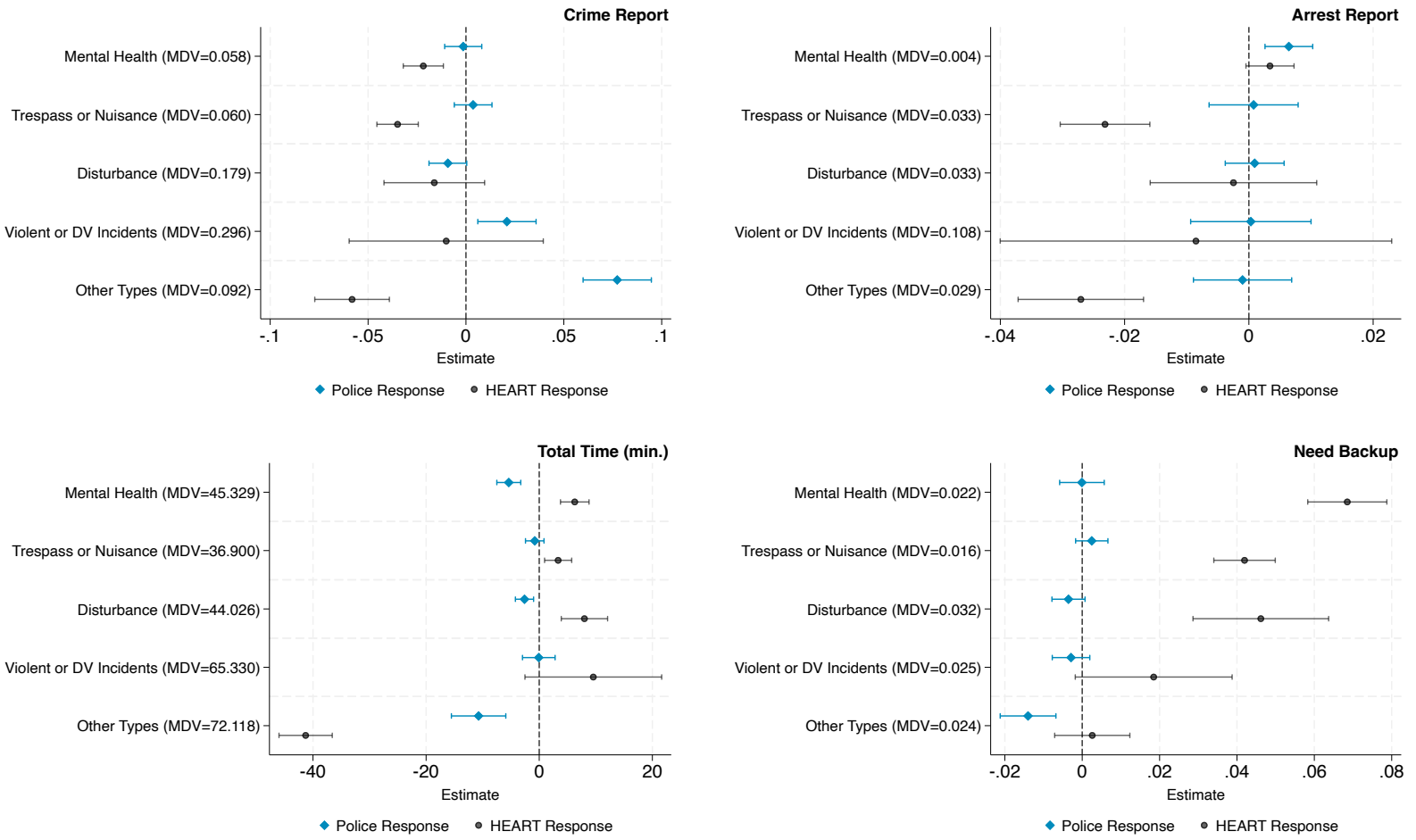
Notes: These figures present the impact of the HEART program on crimes, arrests, dipatch times, and clearing times. They present DiD estimates for police and HEART responses with coefficients and 95% confidence intervals. The mean of the dependent variable for eligible cases in the pre-period is included for reference. Regressions include day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure 6: Impact of HEART Program on Likelihood of a New Call



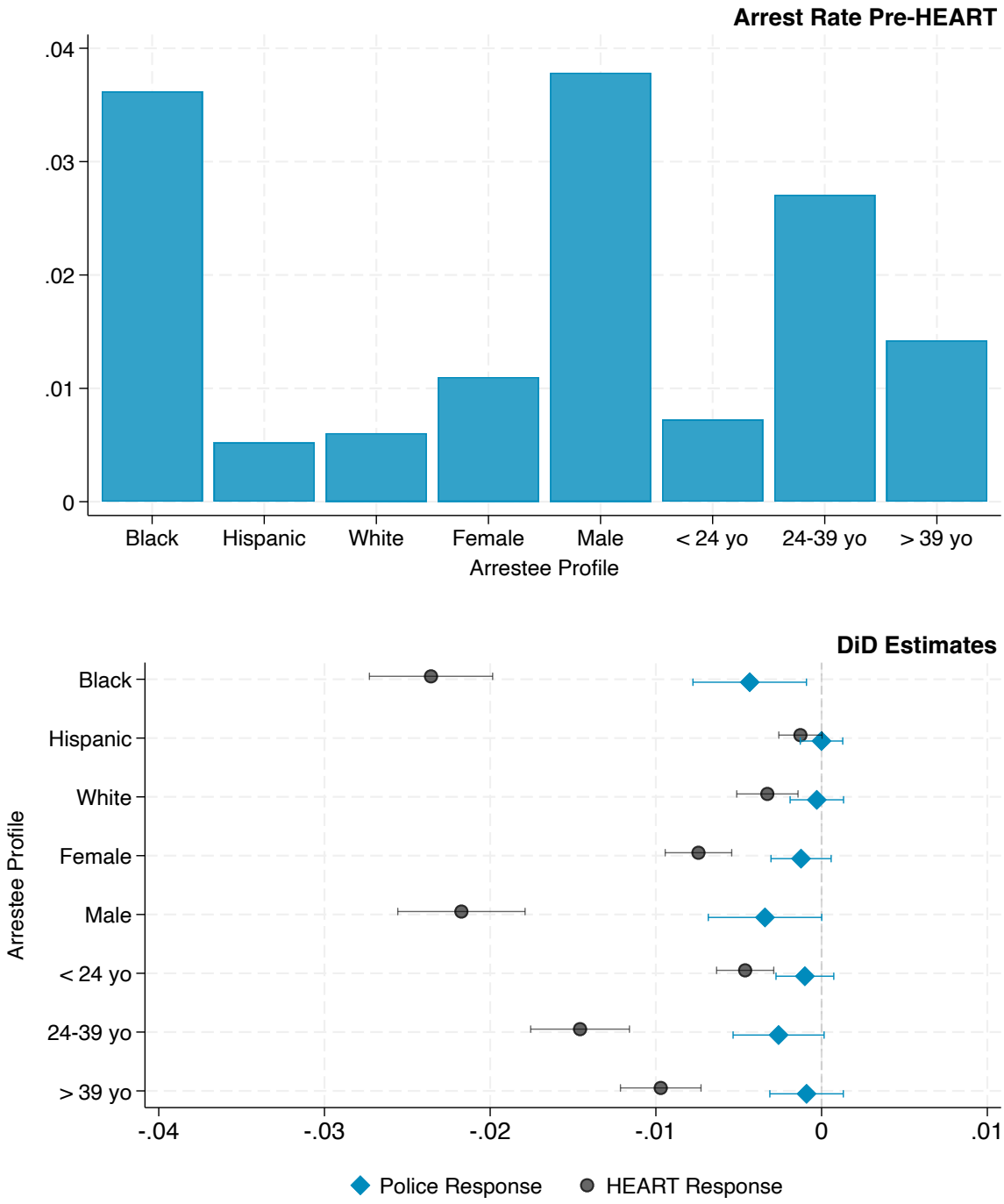
Notes: These figures present the impact of the HEART program on the likelihood of a beneficiary's making a new 911 call in the future. They present DiD estimates for police and HEART responses with coefficients and 95% confidence intervals. The mean of the dependent variable for eligible cases in the pre-period is included for reference. Regressions include day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure 7: Heterogeneity by Nature of Call



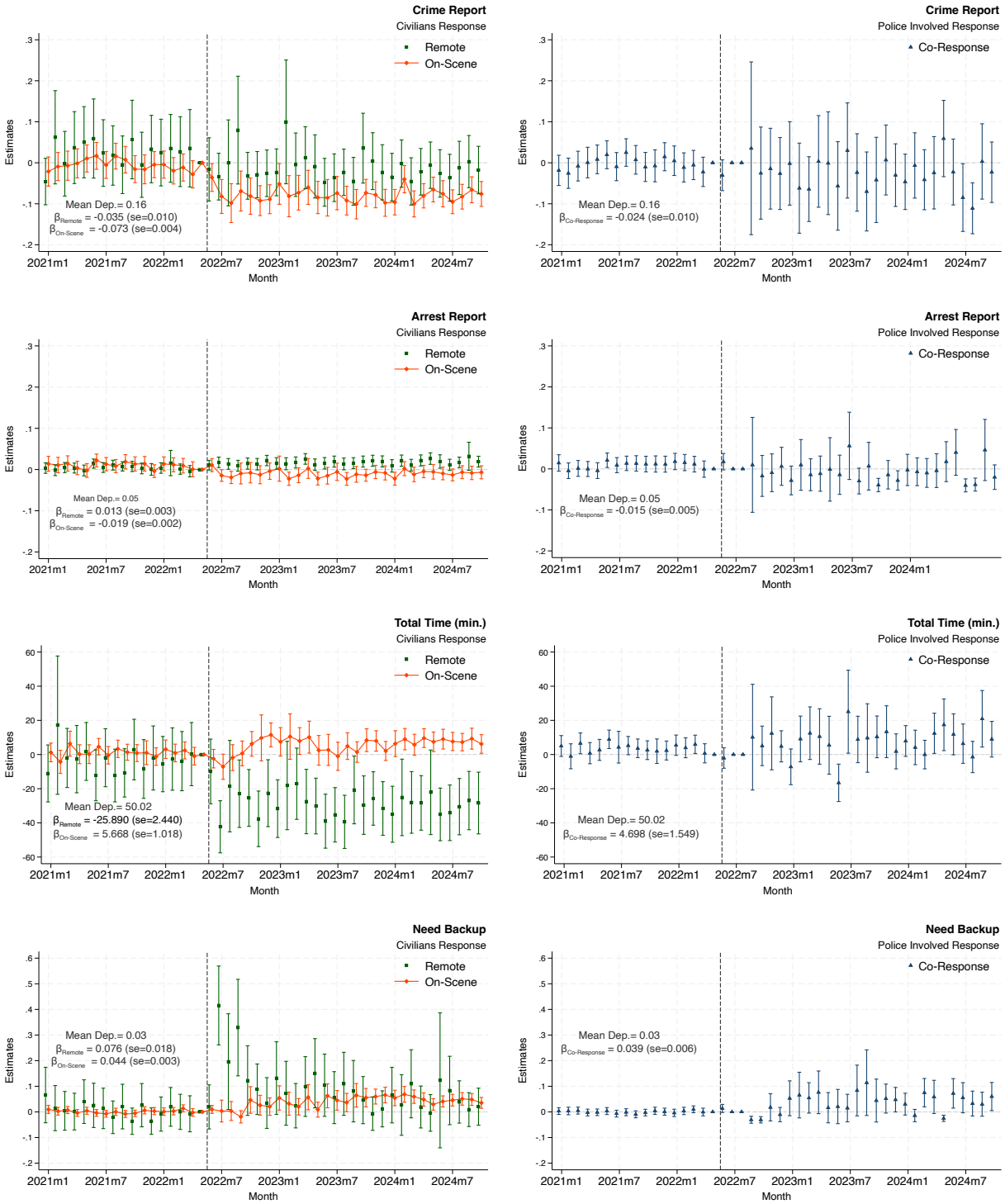
Notes: These figures present the impact of the HEART program on key outcomes, including crimes, arrests, dispatch time, and clearing time, disaggregated by call type: mental health, trespass or nuisance, disturbance, violent or domestic violence incidents, and other types of calls. The "other types" category encompasses calls related to harassment, threats, drug-related incidents, involuntary commitment, and prostitution, among others. We present the mean of the dependent variable (MDV) for each subsample in parentheses. The figures present DiD estimates for police and HEART responses with coefficients and 95% confidence intervals. Regressions include day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure 8: Impact of HEART on Arrestee Profiles



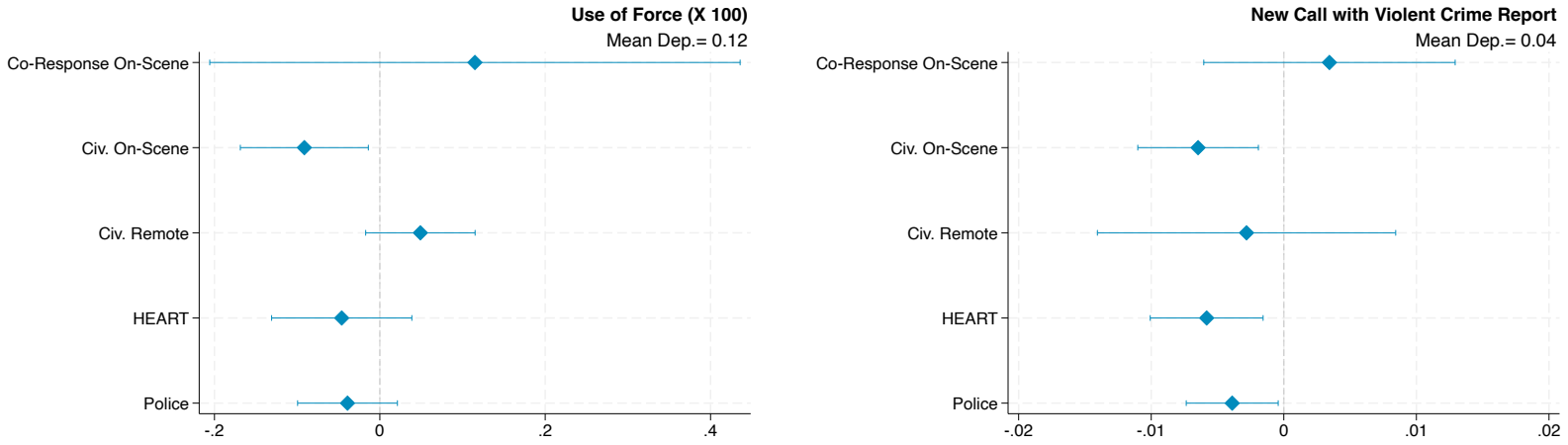
Notes: The figures present the impact of HEART on arrestee profile. The top panel displays pre-HEART arrest rates by demographic group. The bottom panel presents the impact of the HEART program on arrestee characteristics: race, gender, and age. We report the DiD estimates for police and HEART responses with coefficients and 95% confidence intervals. Regressions include day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure 9: Impact on Incident Outcomes by HEART Program



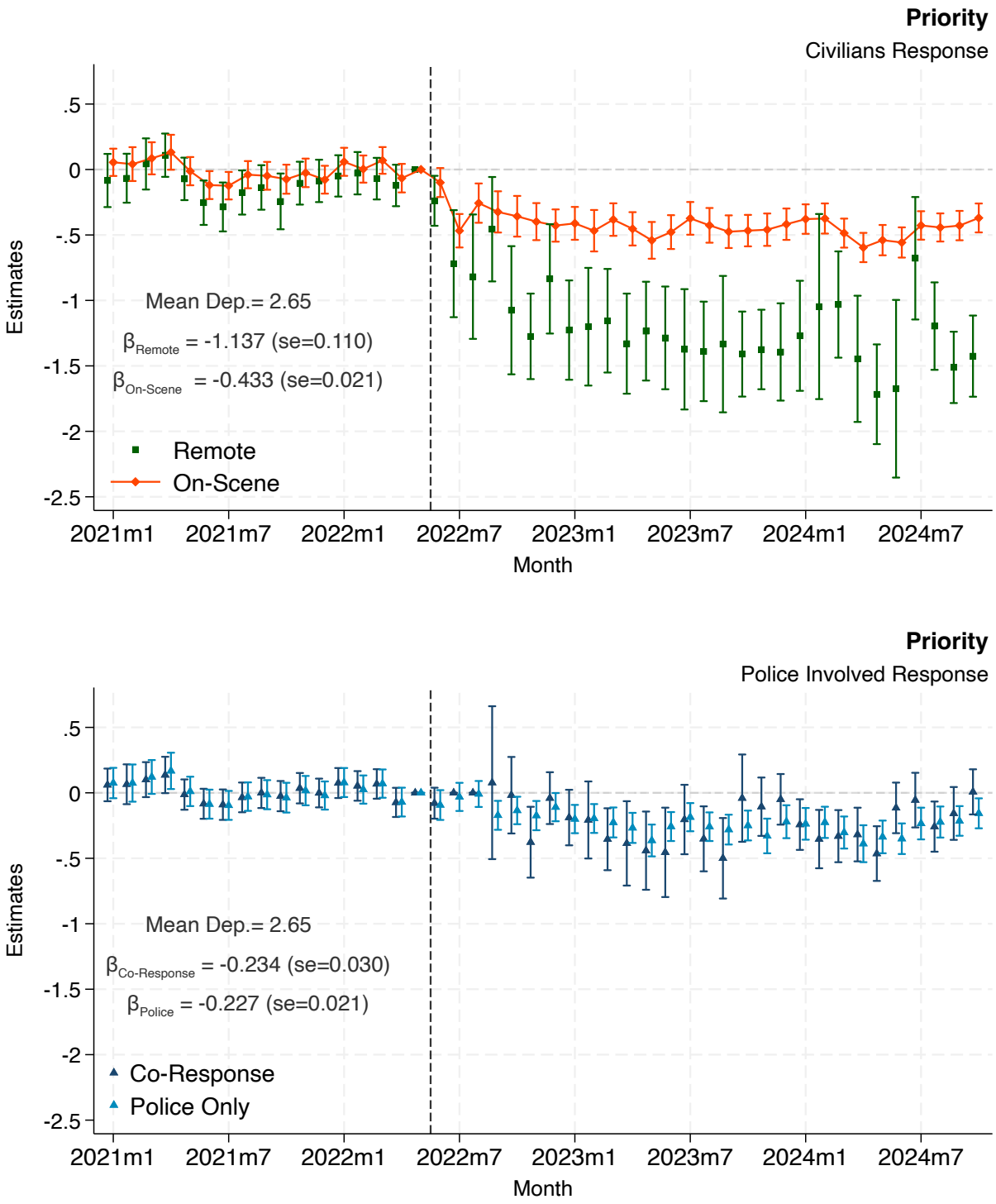
Notes: These figures present the impact of the Crisis Call Diversion (remote civilian response), Community Response Teams (on-scene civilian response), and Co-Response programs on crimes, arrests, dipatch times, and clearing times. Remote civilian response cases involve mental health crises, suicide threats, or crisis calls diverted from police dispatch. The co-scene civilian response cases include both urgent and nonurgent welfare checks, family disturbances, and mental health–related incidents. The figures present DiD estimates for police and HEART responses with coefficients and 95% confidence intervals. The mean of the dependent variable for eligible cases in the pre-period is included for reference. Regressions include day-of-week, hour, call priority, month-year, and beat fixed effects, with standard errors clustered at the address level.

Figure 10: Does HEART Lead to Escalation?



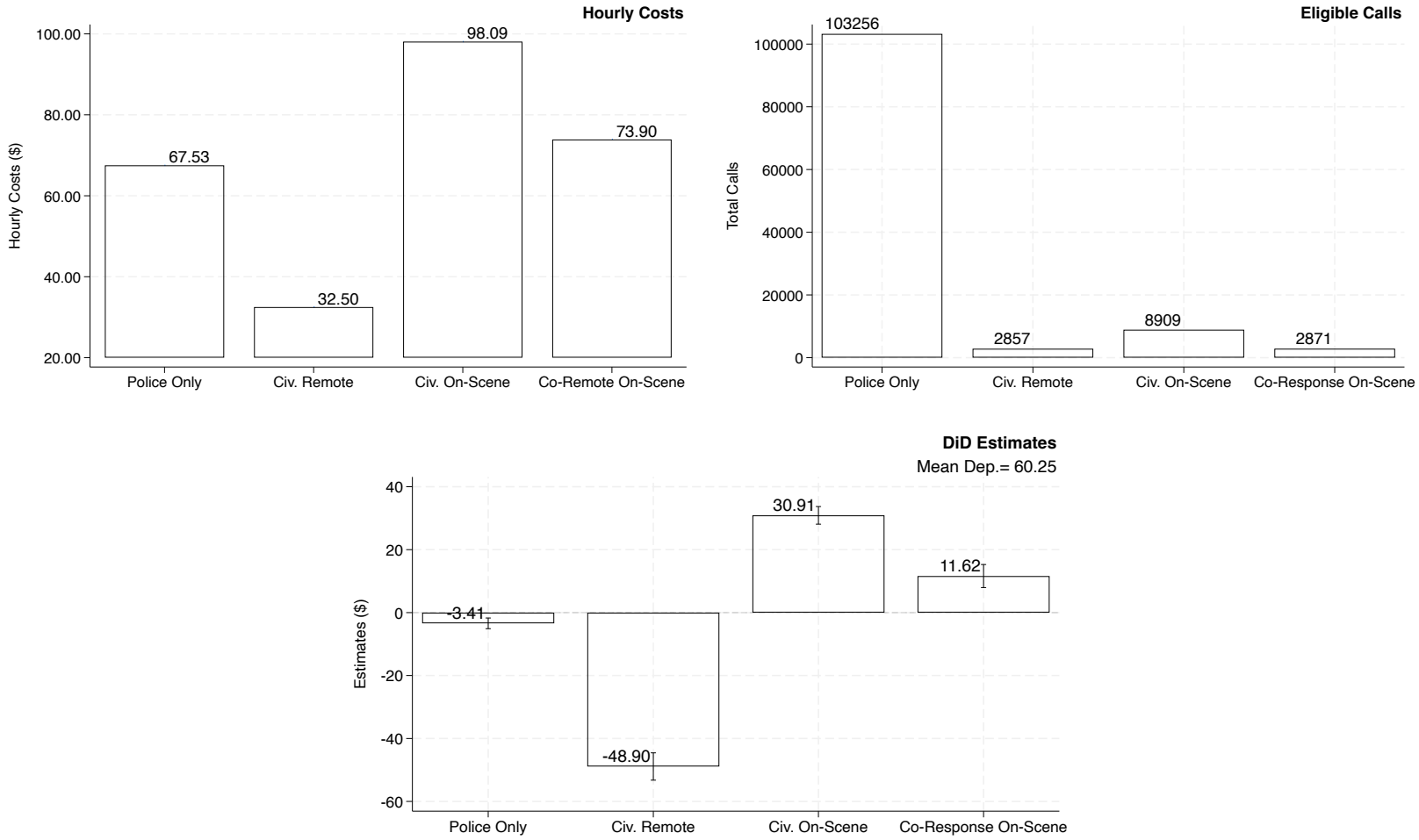
Notes: The figures present the impact of HEART on the likelihood of escalation. We report DiD estimates with 95% confidence intervals, evaluating the effects of police, HEART, Crisis Call Diversion (remote civilian response), Community Response Teams (on-scene civilian response), and Co-Response on the use of force and probability of future calls involving violent crime. The mean of the dependent variable for eligible cases in the pre-period is included for reference. Regressions include day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure 11: Impact on Call Priority by HEART Program



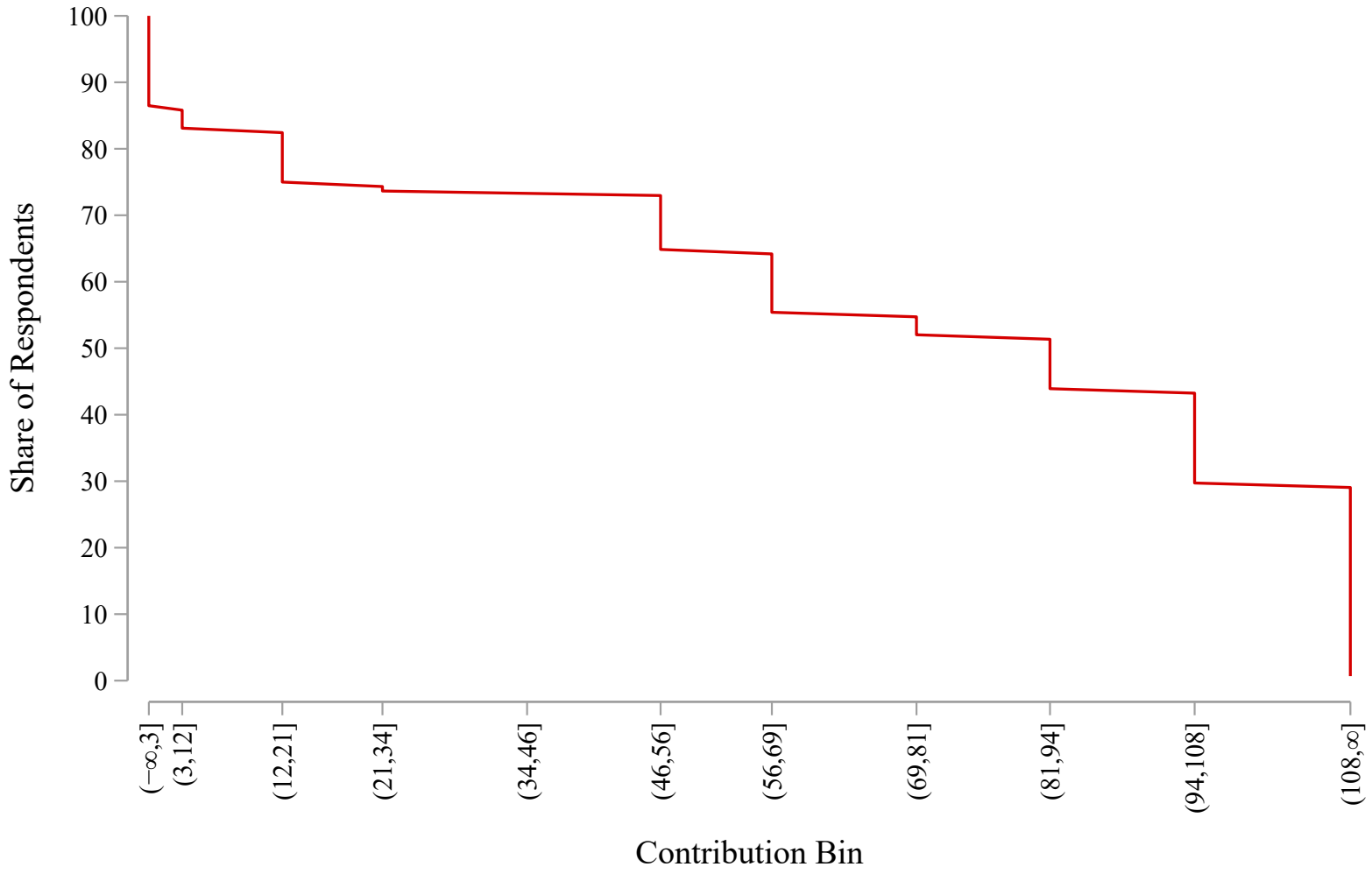
Notes: The figures present the impact of HEART on call priority. Call priority ranges from 0 to 5, where higher numbers correspond to higher priority. We report DiD estimates with 95% confidence intervals, evaluating the effects of police, HEART, Crisis Call Diversion (remote civilian response), Community Response Teams (on-scene civilian response), and Co-Response on the priority score. The mean of the dependent variable for eligible cases in the pre-period is included for reference. Regressions include day-of-week, hour, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure 12: Impact of HEART on Labor Costs



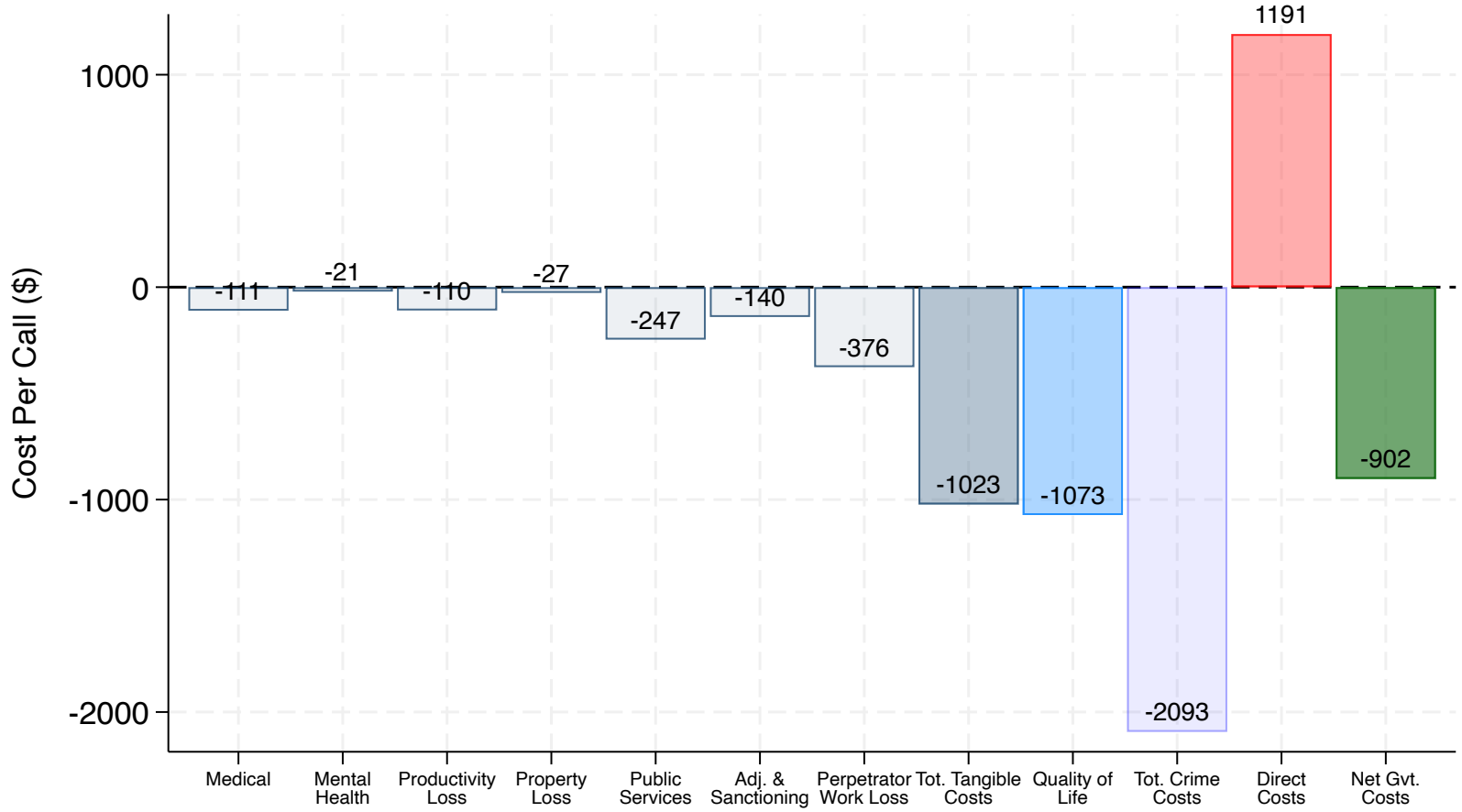
Notes: The top-left figure shows the average hourly wages for police officers and HEART response models—Crisis Call Diversion (remote civilian response), Community Response Teams (on-scene civilian response), and Co-Response programs on the outcome of interest.—calculated as annual salary divided by 52 weeks and 40 hours per week, based on position descriptions from HEART and Durham's public safety agencies. Estimates reflect mean wages within each job category, with police wages reflecting two officers per response, the on-scene civilian response wages covering clinicians, EMTs, and safety specialists, and co-response wages including police paired with mental health professionals. The top-right figure presents the total number of HEART-eligible calls by response type. The bottom figure pesents HEART's impact on labor costs, computed as the product of average hourly wages and response time. DiD estimates with 95% confidence intervals assess the effects of police, HEART, and individual HEART programs on labor costs. The mean of the dependent variable for eligible cases in the pre-period is provided for reference. Regressions control for time and location factors, including day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure 13: WTP Distribution for HEART



Notes: This figure presents the portion of respondents who are willing to pay different amounts for HEART from the contingent valuation survey.

Figure 14: Direct Costs of HEART and its Impact on Fiscal Externality



Notes: This figure illustrates the direct costs of HEART and its impact on fiscal externalities. The bars represent various cost components, with negative values indicating cost savings and positive values representing cost increases. The categories include medical costs, mental health costs, productivity loss, property loss, public services, adjudication and sanctioning, perpetrator work loss, and quality of life impacts. “Total Tangible Costs” comprise the sum of medical costs, mental health costs, productivity loss, property loss, public services, adjudication and sanctioning, and perpetrator work loss. “Total Crime Costs” aggregate these components, while ”Direct Costs” represent HEART’s operational expenses. ”Net Government Costs” reflect the overall fiscal impact, calculated as ”Direct Costs” minus ”Total Crime Costs”. All values are reported in dollars.

Table 1: Summary Statistics

	(1) All Cases	(2) Ineligible Police Resp.	(3) Eligibe Police Resp.	(4) Eligible HEART Resp.	(5) Eligible (4)–(3) Coeff	(6) Eligible (4)–(3) p-value
A) Call Characteristics						
Share Asian	0.039	0.041	0.039	0.035	0.000	0.533
Share Black	0.413	0.408	0.416	0.412	-0.004	0.092
Share Hispanic	0.177	0.173	0.178	0.183	-0.002	0.103
Share Other Race	0.031	0.031	0.031	0.030	0.000	0.323
Unemployment Rate	0.314	0.314	0.313	0.320	-0.000	0.904
Log Rent	10.166	10.191	10.159	10.144	0.011	0.066
Median HH Income (1,000)	67.371	69.090	66.655	67.353	0.800	0.029
B) Crime Outcomes						
Crime Report	0.122	0.071	0.154	0.040	-0.095	0.000
Violent Crime	0.046	0.009	0.067	0.012	-0.026	0.000
Property Crime	0.009	0.009	0.010	0.002	-0.003	0.000
Other Crime	0.077	0.056	0.093	0.028	-0.067	0.000
Arrest Report	0.032	0.017	0.042	0.008	-0.024	0.000
Arrest for Violent Crime	0.016	0.003	0.023	0.003	-0.007	0.000
Arrest for Property Crime	0.003	0.002	0.004	0.001	-0.001	0.000
Arrest for Other Crime	0.022	0.014	0.027	0.006	-0.019	0.000
C) Other Outcomes						
Need Backup	0.022	0.010	0.022	0.058	0.045	0.000
Total Time (min.)	44.961	36.179	48.755	44.089	-2.297	0.022
Dispatch Time (min.)	5.794	5.544	6.011	4.997	-2.540	0.000
Clearing Time (min.)	39.167	30.636	42.744	39.091	0.244	0.772
Use of Force ($\times$100)	0.084	0.051	0.103	0.048	-0.008	0.824
Any New Call	0.897	0.847	0.921	0.875	-0.010	0.017
New Call with Violence	0.026	0.018	0.030	0.020	-0.002	0.132
Observations	161054	43161	103256	14637	117893	117893

Notes: The table presents summary statistics for HEART-ineligible, police response, and HEART response cases. Column (1) reports the mean for the full sample. Columns (2)-–(4) show the means by case type: ineligible police responses, eligible police responses, and eligible HEART responses. Column (5) presents the difference in means between HEART and police responses among eligible cases, while Column (6) reports the p-value for the test of equality between these groups. The regression includes day-of-week, hour, priority, month–year, and beat fixed effects. Standard errors are clustered at the address level.

Table 2: Impact of HEART on Various Outcomes
Alternatives Research Designs

	(1) Mean Pre-HEART	(2) Main Analysis Estimate	(3) Main Analysis SE	(4) Triple Diff. Estimate	(5) Triple Diff. SE	(6) TSLS Estimate	(7) TSLS SE	(8) Fuzzy DID Estimate	(9) Fuzzy DID SE
A) Crime									
Crime Report	0.161	-0.092	0.005	-0.094	0.005	-0.079	0.024	-0.080	0.011
Violent Crime	0.078	-0.046	0.002	-0.039	0.002	-0.086	0.014	-0.032	0.007
Property Crime	0.013	-0.010	0.001	-0.008	0.001	-0.032	0.008	-0.008	0.002
Other Crime	0.093	-0.051	0.004	-0.059	0.004	-0.021	0.020	-0.049	0.010
B) Arrest									
Arrest Report	0.048	-0.028	0.002	-0.029	0.002	-0.057	0.013	-0.022	0.003
Arrest for Violent Crime	0.028	-0.016	0.001	-0.013	0.001	-0.035	0.008	-0.006	0.001
Arrest for Property Crime	0.006	-0.005	0.001	-0.004	0.001	-0.020	0.004	-0.004	0.001
Arrest for Other Crime	0.029	-0.018	0.002	-0.020	0.002	-0.032	0.011	-0.018	0.003
C) Other Outcomes									
Need Back-Up	0.025	0.041	0.003	0.040	0.003	0.022	0.010	0.037	0.003
Total Time (min.)	50.017	-5.894	1.030	-5.380	1.031	-21.556	5.117	-7.451	1.936
Dispatch Time (min.)	4.941	-1.828	0.212	-2.136	0.224	-1.105	1.079	-3.582	1.630
Clearing Time (min.)	45.076	-4.066	1.006	-3.245	0.988	-20.450	4.974	-3.869	1.784
Use of Force ($\times$ 100)	0.130	-0.046	0.043	-0.047	0.043	-0.266	0.209	-0.027	0.048
Any New Call	0.960	0.021	0.006	0.018	0.005	0.182	0.029	-0.014	0.009
New Call with Violence	0.034	-0.006	0.002	-0.006	0.002	-0.029	0.012	-0.003	0.003

Notes: The table summarizes the estimates of the impact of HEART on various outcomes across multiple research designs. Column (1) presents the mean of each outcome in the pre-HEART period. Even-numbered columns report the estimated effects, while odd-numbered columns display the corresponding standard errors. The table includes results from the main DiD specification, triple differences, TSLS, and fuzzy DiD. All regressions account for day-of-week, hour, priority, month–year, and beat fixed effects. Standard errors are clustered at the address level.

Contents

1 Introduction **1**

2 Background on Emergency Responses **6**
2.1 Crisis Response Programs in the US 6
2.2 Durham's HEART 7

3 Data and Descriptives Facts **9**
3.1 Data Construction 9
3.2 Descriptives 10

4 Empirical Strategy **11**

5 Police vs. HEART Responses **13**
5.1 Main Results 13
5.1.1 Impact of HEART on Incident Outcomes 13
5.1.2 Impact on Repeat 911 Calls 14
5.1.3 Heterogeneity by Nature of the Call 15
5.2 Arrestee Profiles and HEART 15
5.3 Robustness 16
5.3.1 DiD Assumptions 16
5.3.2 Alternative Designs 17

6 Impact of Each Program **17**
6.1 Remote vs. In-Person and Police-Involved Responses 17
6.2 Does HEART Lead to Escalation? 18
6.3 Does HEART Improve Trust in 911? 19
6.4 Labor Costs 20
6.5 Robustness 21

7 Welfare Implications **21**
7.1 Government Costs 21
7.2 Measuring Public WTP Using Contingent Valuation Survey 21
7.3 Measuring Cost-Effectiveness with MVPF 23

8 Conclusion **25**

A Data **1**

B Additional Specifications **1**
B.1 Impact of HEART from TSLS 1
B.2 Triple-Difference Design 2

Supplementary Materials

A Data

B Additional Specifications

B.1 Impact of HEART from TSLS

Estimation Strategy As a complementary analysis, we follow Ouss and Stevenson (2023) and estimate the effect of a HEART response on various outcomes with TSLS by taking advantage of the sharp introduction of the HEART program and its targeted eligibility criteria, using the variation in the policy's impact on eligible calls as an instrument for HEART responses. We recover the TSLS estimates by estimating the following equations:

$$HEART_{it} = \alpha_F + Post_t \cdot Eligible_i \cdot \beta^F + Eligible_i \cdot \lambda_F + \gamma_t^F + X_{ic}' \cdot \delta_F + \epsilon_{ict}^F \tag{5}$$

$$Y_{ict} = \alpha_H + HEART_{it} \cdot \beta^H + Eligible_i \cdot \lambda_H + \gamma_t^H + X_{ic}' \cdot \delta_H + \epsilon_{ict}^H \tag{6}$$

where the variable $Eligible_i$ indicates whether a call is eligible for a HEART response and $Post_t$ equals one for calls occurring after the HEART program's introduction in June 2022 and zero otherwise. The interaction term, $Post_t \cdot Eligible_i$, serves as our instrument, leveraging the sharp introduction of HEART and the eligibility criteria. The key identifying assumption for the first stage under DiD is that in the absence of HEART's sharp introduction, trends in outcomes for HEART-eligible and HEART-ineligible calls with a police response would have remained parallel.

Following Imbens and Angrist (1994), the key identifying assumptions for our design are relevance, independence, monotonicity and the exclusion restriction. Relevance requires that the instrument significantly predict the likelihood of a HEART response, meaning that eligible calls post-introduction must be likelier to receive the intervention. Independence assumes that the potential outcomes are independent of the HEART assignment, conditional on the instrument, indicating that trends in outcomes would be similar for eligible and ineligible calls without the program. The exclusion restriction requires that the instrument affect the outcome only through its impact on treatment probability. Finally, monotonicity requires that the instrument consistently increase the likelihood of receiving a HEART response.

Relevance and Balance Tests Figure A.10 presents the first-stage regression results, estimating the probability of receipt of a HEART response over time. The vertical dashed line marks the program's introduction, after which the likelihood of a HEART response for eligible calls rises sharply and remains elevated. The coefficient on the interaction between the post-introduction and eligibil-

ity indicators is 0.150 ($SE = 0.004$), with a large F-statistic, confirming a strong first stage. Prior to HEART's implementation, the trends in outcomes for eligible and ineligible calls remained flat and parallel, supporting the identification strategy. The sharp post-introduction divergence indicates the program's substantial impact on HEART response rates, reinforcing the strength of the instrument.

Table A.1 assesses the validity of the independence assumption by testing whether the instrument, $Post \cdot Eligible$, is conditionally independent of call characteristics. A valid instrument should not systematically correlate with observable covariates. Column (1) examines the relationship between HEART responses and call characteristics, showing that HEART responses are slightly less likely in areas with a larger Asian population share but exhibit no strong patterns otherwise. Column (2) tests the instrument, $Post \cdot Eligible$, and finds that its coefficients are small and mostly statistically insignificant, suggesting that the instrument does not predict systematic differences in observable characteristics. The result from the joint F-test for HEART responses (Column (1)) is marginally significant ($p = 0.067$), indicating some correlation with call characteristics, whereas the result from the joint F-test for the instrument (Column (2)) is insignificant ($p = 0.138$), supporting the assumption that the instrument is balanced.

B.2 Triple-Difference Design

Estimation Strategy We analyze the impact of police and HEART responses on eligible cases using a triple-difference design (DDD), where we estimate the following specification:

$$Y_{ict} = \alpha + HEART_{it} \cdot \beta^H + Police^H_{it} \cdot \beta^P + Eligible_i \cdot \lambda + \gamma_t + X'_{ic} \cdot \delta + \epsilon_{ict} \tag{7}$$

where $HEART_{it}$ and $Police^H_{it}$ equals to one if there is a HEART or a police response to eligible cases, and zero otherwise.[1] The parameters of interest are β^H and β^P which capture the change in outcome for HEART and police responses in HEART eligible cases after the deployment of the program relative to ineligible cases.

[1] $HEART_{it} = Post_t \cdot HEART_i \cdot Eligible_i$ and $Police^H_{it} = Post_t \cdot (1 - HEART_i) \cdot Eligible_i$

Additional Figures and Tables

Table A.1: Balance Tests

	(1) HEART	(2) Post×Eligible
Sh. Asian	-0.00697	-0.0558**
	(0.0265)	(0.0244)
Sh. Black	-0.0188	-0.00231
	(0.0120)	(0.00887)
Sh. Hispanic	-0.0267	0.00766
	(0.0186)	(0.0131)
Sh. Other Race	0.0397	-0.0238
	(0.0418)	(0.0372)
Unemployment Rate	0.000609	-0.00632
	(0.0181)	(0.00924)
Median HH Income (1,000)	-0.0000568	0.0000331
	(0.0000810)	(0.0000660)
Log Rent	0.00451	0.00159
	(0.00603)	(0.00491)
Controls	Yes	Yes
Joint F-test	1.619	1.499
p-value joint F-test	0.125	0.162
Observations	161054	161054

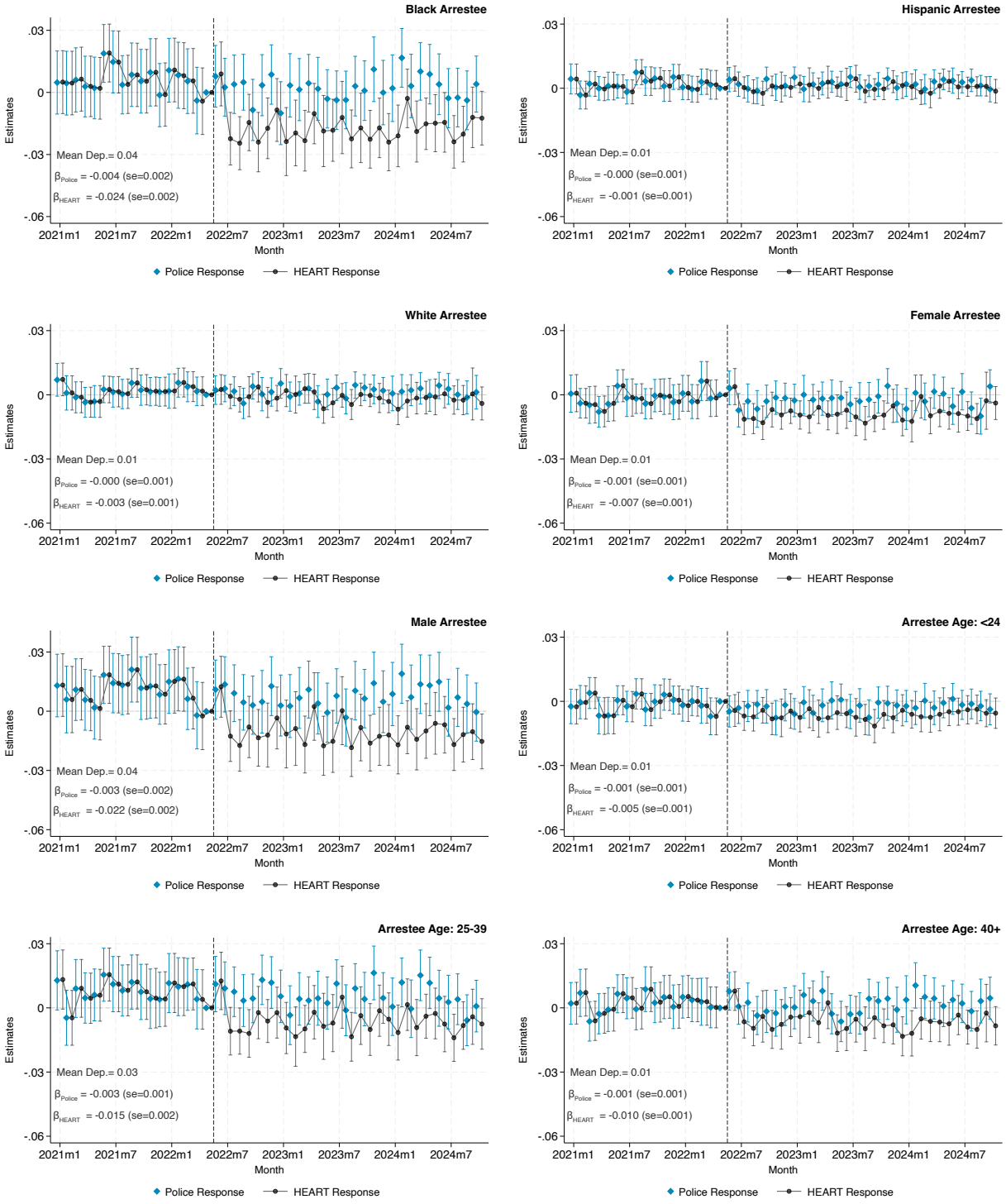
Notes: This table reports reduced-form results from OLS regressions testing for balance. Column (1) examines the correlation between call characteristics and HEART responses, while column (2) focuses on the interaction of Post and Eligible. All regressions control for day of the week, hour, priority, month-year, and beat fixed effects. The p-values and F-statistics at the bottom test the joint significance of the variables listed in each row. Standard errors are clustered at the address level. Significant at 10%; significant at 5%; significant at 1%.

Table A.2: Costs per Crime by Cost Category from Miller et al. (2021)

Crime	Medical	Mental Health	Productivity	Property Loss	Public Services	Adjudication & Sanctioning	Perpetrator Work Loss	Subtotal: Tangible Costs	Quality of Life	Total Cost
Rape	$3,333	$6,504	$7,178	$176	$901	$44,660	$18,409	$81,161	$319,632	$400,793
Other Sex Offenses	$706	$1,580	$1,760	$68	$51	$328	$135	$4,628	$82,507	$87,135
Robbery	$1,959	$196	$4,639	$1,285	$1,321	$13,784	$5,928	$29,112	$14,656	$43,768
Assult	$2,090	$403	$2,292	$79	$4,315	$6,172	$2,286	$17,637	$21,149	$38,786
Arson	$2,647	$45	$3,389	$19,519	$4,002	$2,596	$505	$32,703	$6,430	$39,133
Drving Under the Influence	$3,719	$432	$17,022	$7,848	$78	$1,088	$107	$30,294	$53,449	$83,743
Burglary	$0	$0	$39	$2,882	$582	$935	$931	$5,369	$0	$5,369
Larceny/Theft	$0	$0	$31	$1,052	$901	$2,570	$226	$4,780	$0	$4,780
Motor Vehicle Theft	$0	$0	$118	$7,219	$715	$1,964	$767	$10,783	$0	$10,783
Fraud	$0	$0	$57	$1,854	$73	$52	$16	$2,052	$0	$2,052
Stolen Property Offenses	$0	$0	$0	$0	$1,321	$5,385	$1,570	$8,276	$0	$8,276
Vandalism	$0	$0	$0	$390	$23	$688	$248	$1,349	$0	$1,349
Weapon Law Violations	$0	$0	$0	$0	$79	$2,573	$1,073	$3,725	$0	$3,725
Prostitution/Pandering	$0	$0	$0	$0	$79	$257	$108	$444	$0	$444
Drug Offenses	$0	$0	$0	$0	$5,046	$3,599	$1,502	$10,147	$0	$10,147
Gambling	$0	$0	$0	$0	$79	$257	$108	$444	$0	$444
Liquor Law Violations	$0	$0	$0	$0	$79	$1,228	$512	$1,819	$0	$1,819
Drunkenness	$0	$0	$0	$0	$79	$1,228	$512	$1,819	$0	$1,819
Disorderly Conduct	$0	$0	$0	$0	$79	$1,228	$512	$1,819	$0	$1,819
Vagrancy	$0	$0	$0	$0	$79	$1,228	$512	$1,819	$0	$1,819
Curfew/Loitering Violations	$0	$0	$0	$0	$79	$1,228	$512	$1,819	$0	$1,819
All Other Non-traffic Offenses	$0	$0	$0	$0	$79	$257	$165	$501	$0	$501

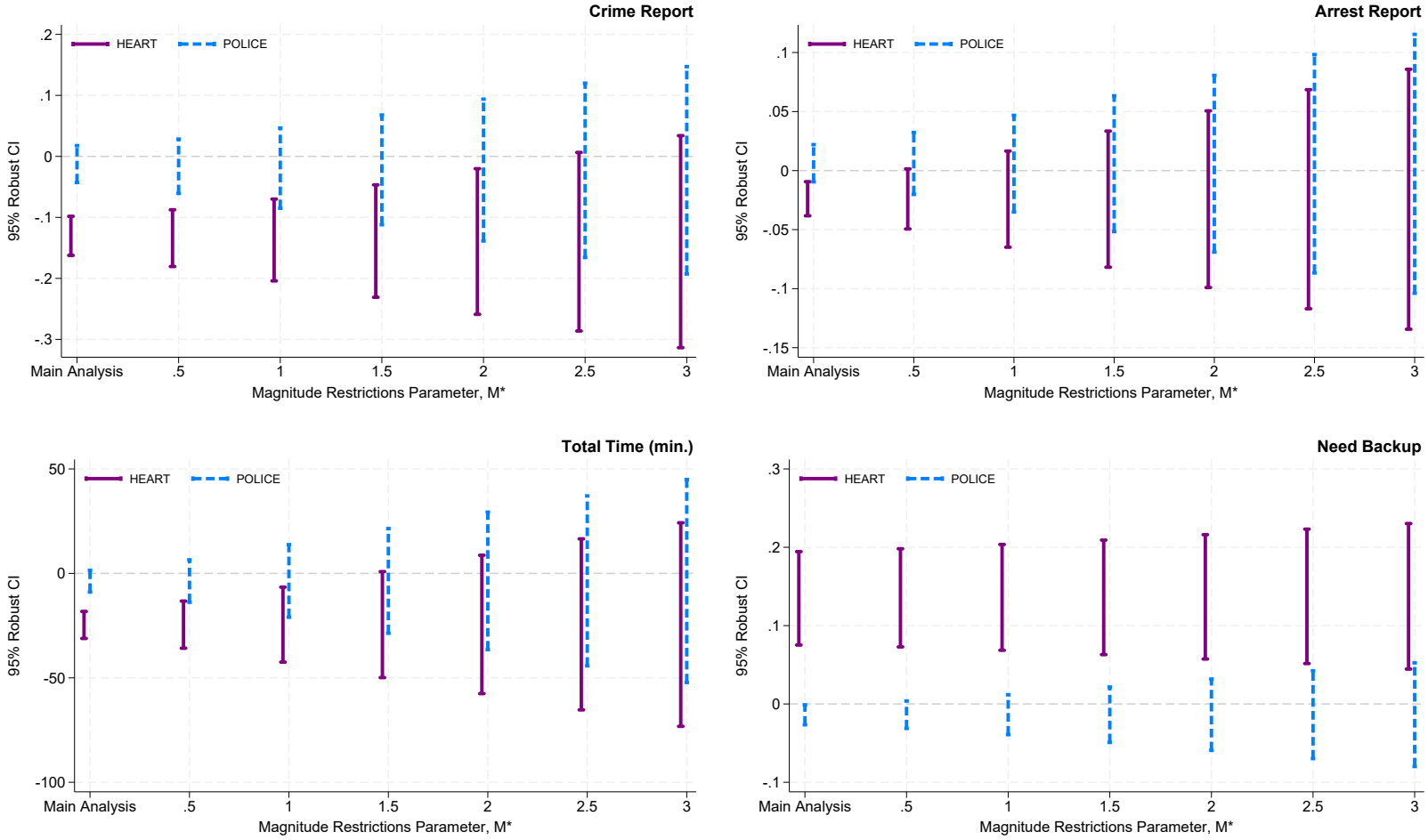
Notes: This table displays the per-crime cost estimates across various cost categories, including incidents that were not reported to the police in the United States. All estimates are in 2017 dollars and are sourced from Miller et al. (2021) . Public services encompass police, fire, EMS, and victim services. The violent crime categories include rape, other sexual assault, robbery, assault, intimate partner violence, child maltreatment, and arson .

Figure A.1: Impact of the HEART Program on Arrestees' Characteristics



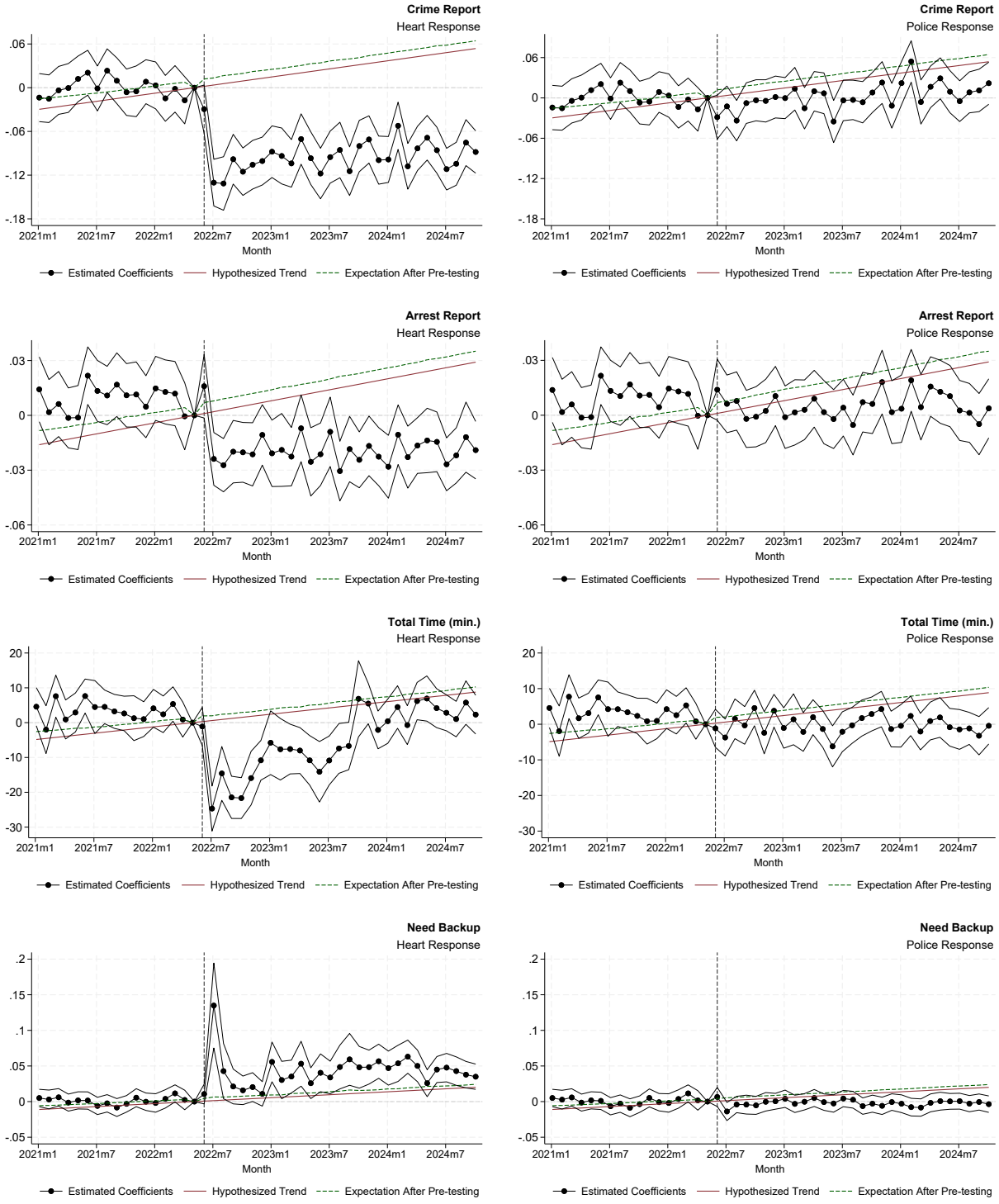
Notes: These figures present the impact of the HEART program on arrestees' characteristics. It presents difference-in-differences (DiD) estimates for police and HEART responses, with coefficients and 95% confidence intervals. The mean of the dependent variable for eligible cases in the pre-period is included for reference. Regressions include day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure A.2: Bounds on the Relative Magnitudes for DiD



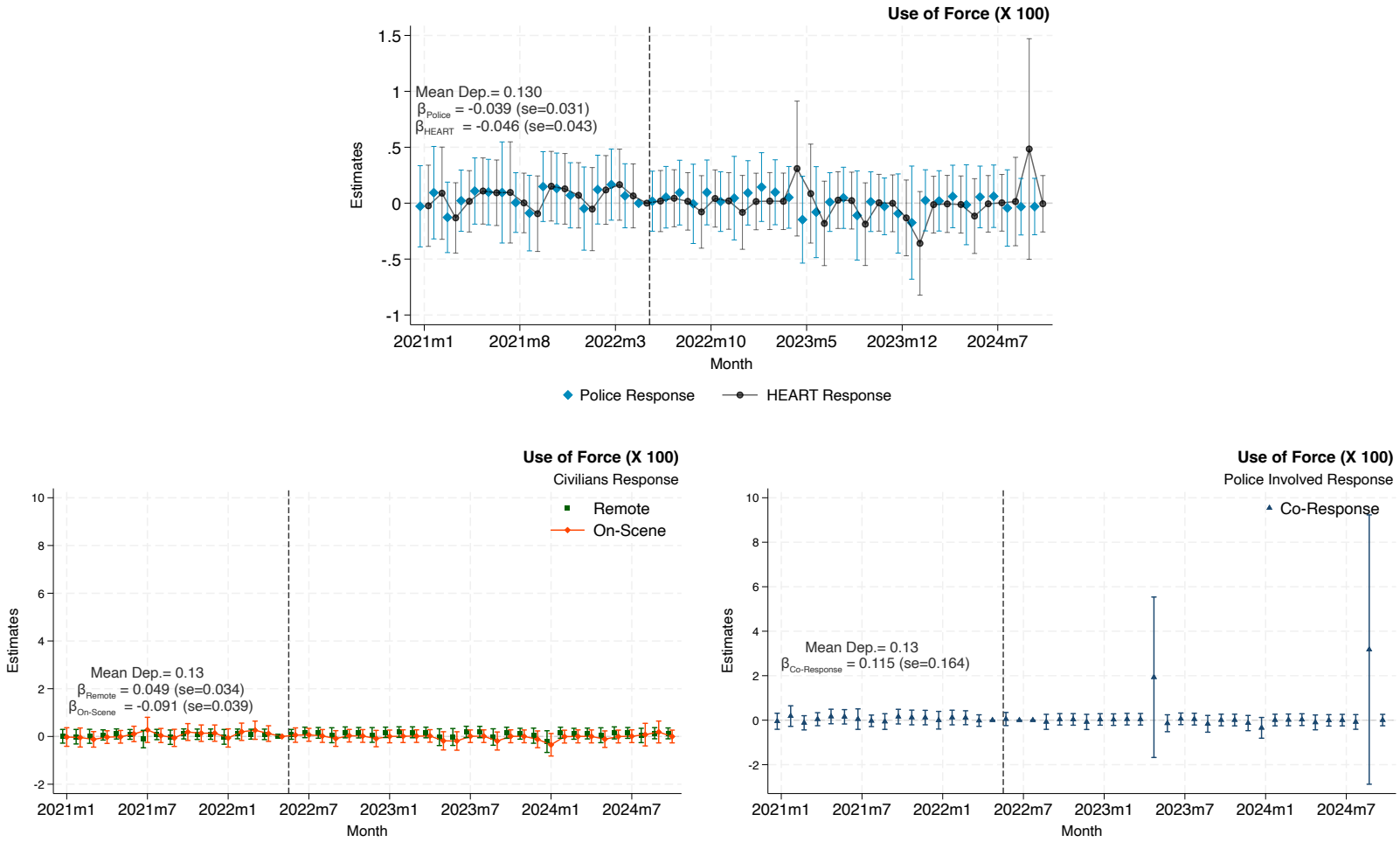
Notes: These figures present various specification checks on the impact of the HEART on various outcomes. The figure presents the bounds on relative magnitudes associated with a 95% confidence interval from Rambachan and Roth (2023).

Figure A.3: Event Study and Hypothesized Trends



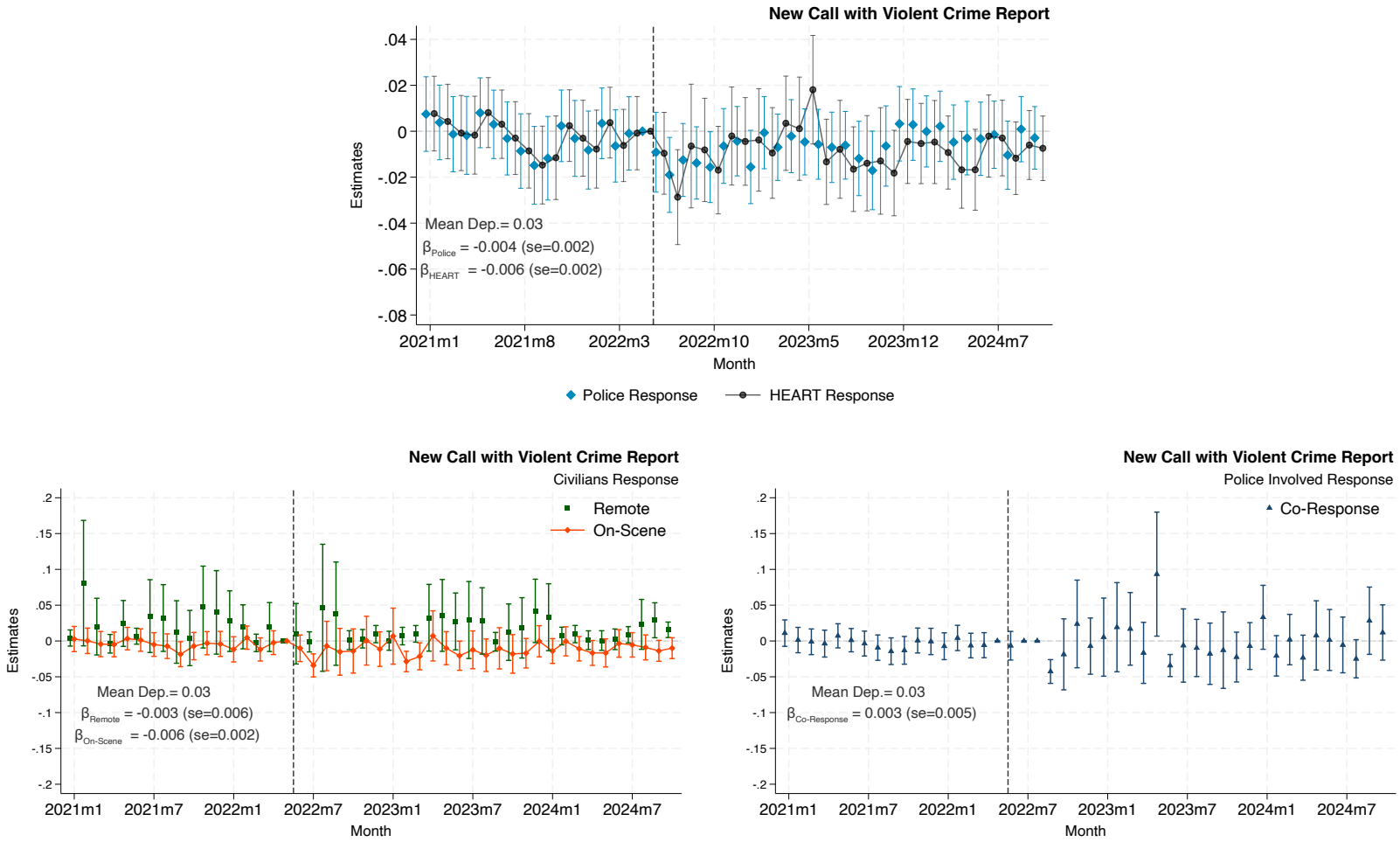
Notes: These figures present various specification checks on the impact of the HEART on various outcomes.The figures plot potential violations of parallel trends based on Roth (2022). We report the event study coefficients from Figure 4 and 95% confidence intervals, with standard errors clustered at the address level. The solid line indicates the hypothesized linear deviation from parallel trends with 80% power. The dashed line shows the expected values of event study coefficients if the deviation existed but was undetectable using conventional methods.

Figure A.4: Impact of HEART Program on Police Use of Force



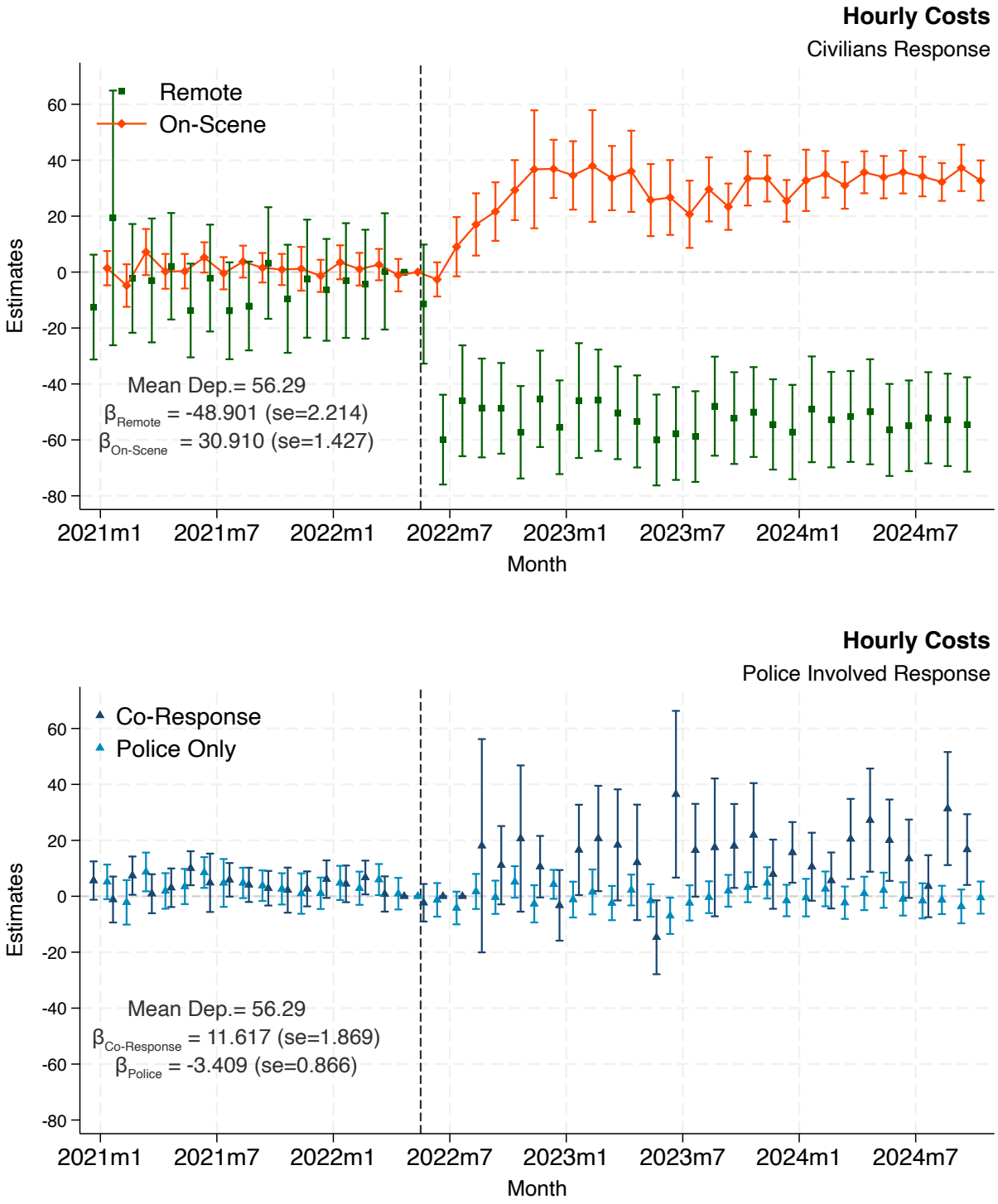
Notes: The top panel present the impact of the HEART program on use of force (X 100). The bottom panel focus on each program. These figures present the impact of the Crisis Call Diversion (remote civilian response), Community Response Teams (on-scene civilian response), and Co-Response programs on use of force (X 100). Remote civilian responses involve mental health crises, suicide threats, or crisis calls diverted from police dispatch. The on-scene civilian response cases include both urgent and non-urgent welfare checks, family disturbances, and mental health-related incidents. It presents difference-in-differences (DiD) estimates for police and HEART responses, with coefficients and 95% confidence intervals. The mean of the dependent variable for eligible cases in the pre-period is included for reference. Regressions include day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure A.5: Impact of HEART Program on New Call for Violent Crimes



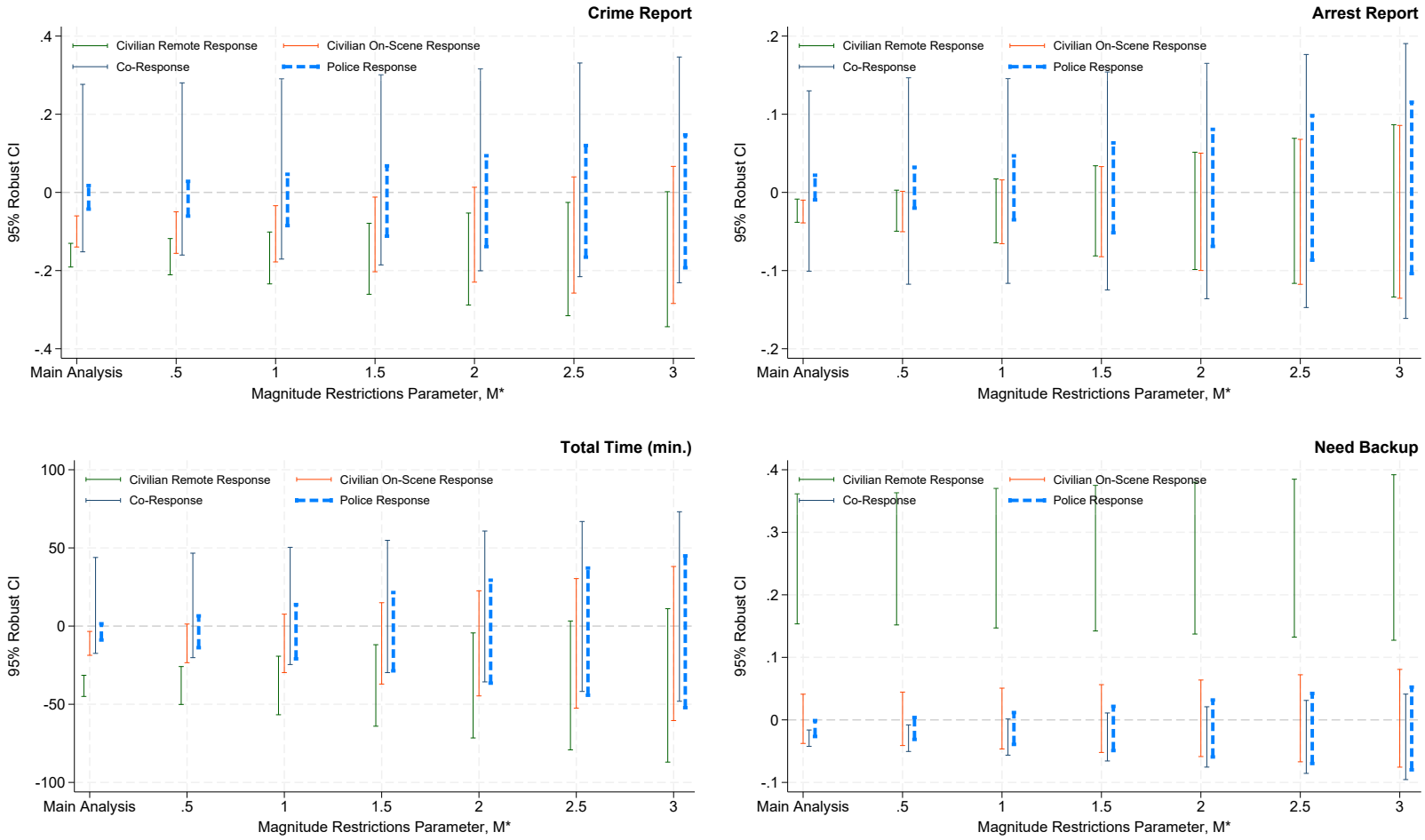
Notes: The top panel present the impact of the HEART program on the likelihood of a new call for violent crime. The bottom panel focus on each program. These figures present the impact of the Crisis Call Diversion (remote civilian response), Community Response Teams (on-scene civilian response), and Co-Response programs on the outcome of interest. Remote civilian responses involve mental health crises, suicide threats, or crisis calls diverted from police dispatch. The on-scene civilian response cases include both urgent and non-urgent welfare checks, family disturbances, and mental health-related incidents. It presents difference-in-differences (DiD) estimates for police and HEART responses, with coefficients and 95% confidence intervals. The mean of the dependent variable for eligible cases in the pre-period is included for reference. Regressions include day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure A.6: Impact of HEART on Labor Costs



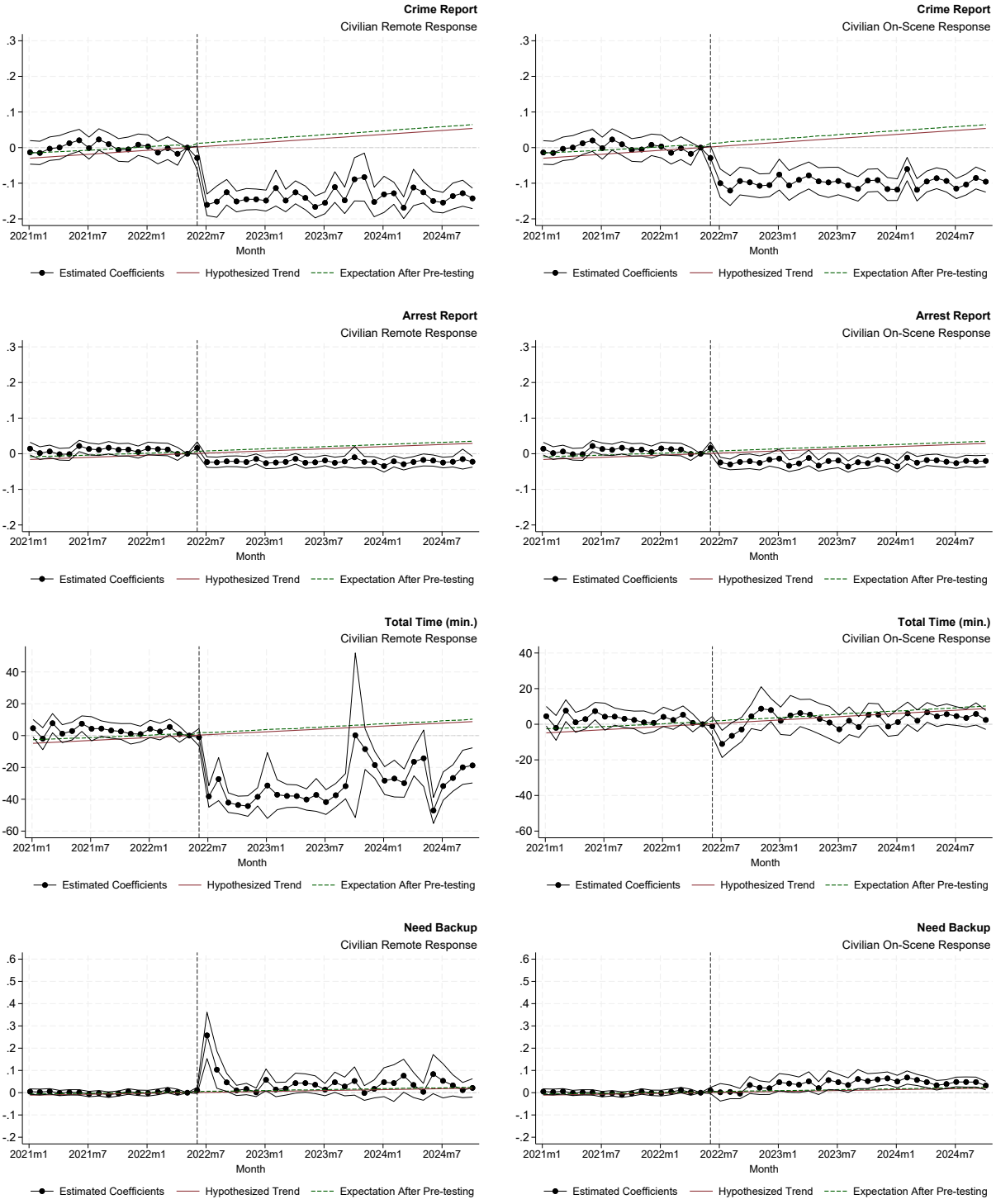
Notes: The figures present the impact of HEART on labor costs, which are calculated by multiplying the average hourly wage of each response type by the time required to handle the calls. We report difference-in-differences (DiD) estimates with 95% confidence intervals, evaluating the effects of police, Crisis Call Diversion (remote civilian response), Community Response Teams (on-scene civilian response), and Co-Response programs , or police only on labor costs. The mean of the dependent variable for eligible cases in the pre-period is included for reference. Regressions include day-of-week, hour, call priority, month–year, and beat fixed effects, with standard errors clustered at the address level.

Figure A.7: Bounds on the Relative Magnitudes for DiD for Each HEART Program



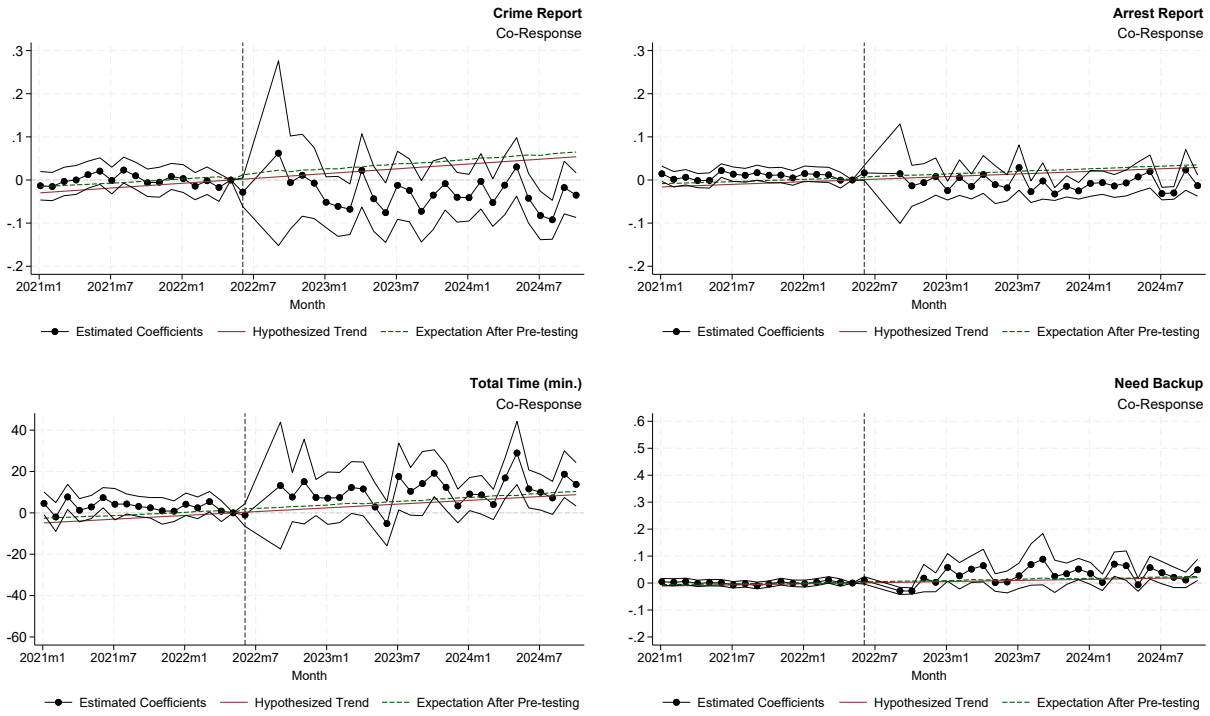
Notes: These figures present various specification checks on the impact of each HEART program on various outcomes. The figure presents the bounds on relative magnitudes associated with a 95% confidence interval from Rambachan and Roth (2023).

Figure A.8: Event Study and Hypothesized Trends for Civilian Responses



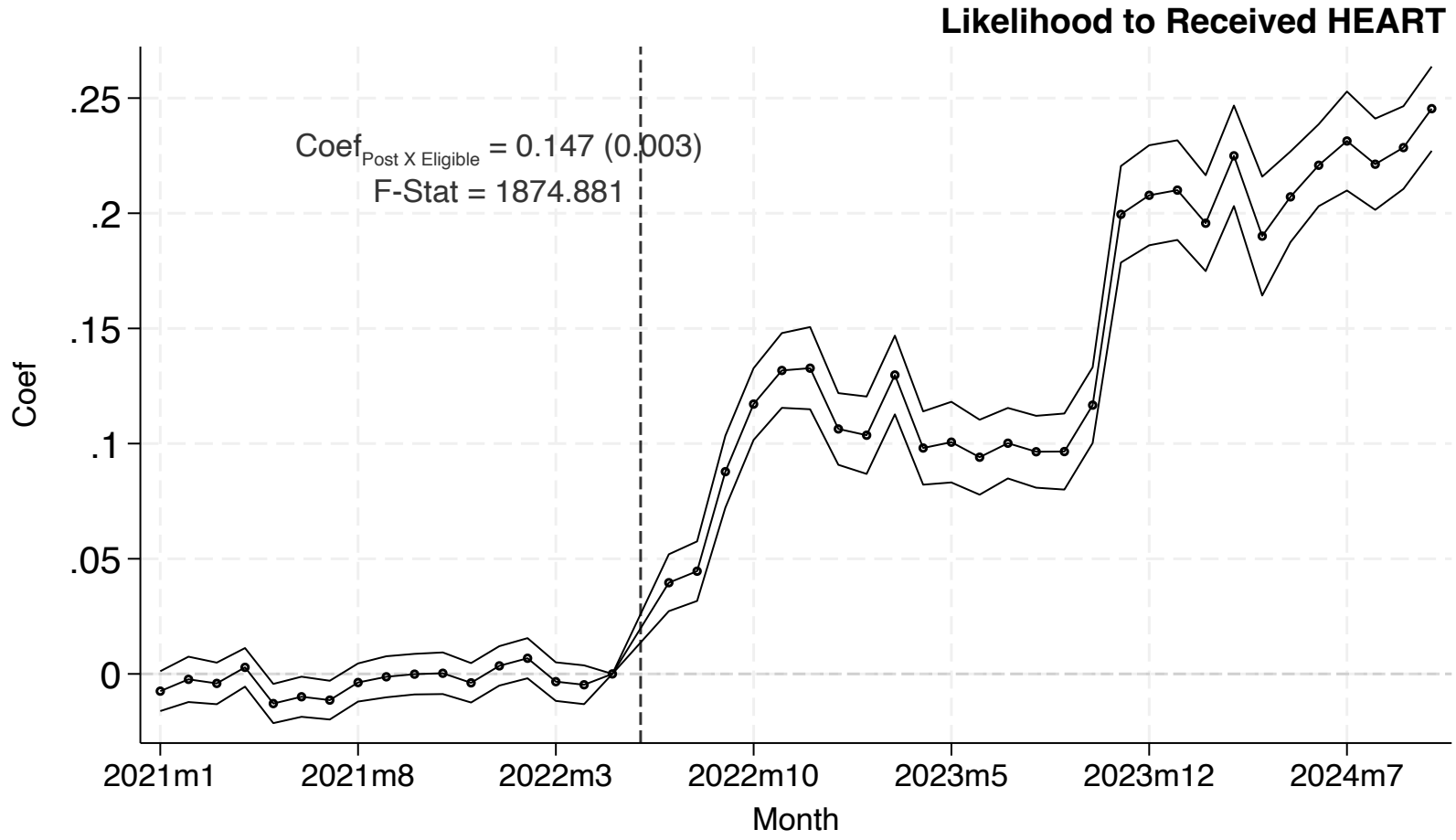
Notes: These figures present various specification checks on the impact of civilians responses programs on various outcomes.The figures plot potential violations of parallel trends based on Roth (2022). We report the event study coefficients from Figure 4 and 95% confidence intervals, with standard errors clustered at the address level. The solid line indicates the hypothesized linear deviation from parallel trends with 80% power. The dashed line shows the expected values of event study coefficients if the deviation existed but was undetectable using conventional methods.

Figure A.9: Event Study and Hypothesized Trends for Co-Response



Notes: These figures present various specification checks on the impact of co-responses on various outcomes.The figures plot potential violations of parallel trends based on Roth (2022). We report the event study coefficients from Figure 4 and 95% confidence intervals, with standard errors clustered at the address level. The solid line indicates the hypothesized linear deviation from parallel trends with 80% power. The dashed line shows the expected values of event study coefficients if the deviation existed but was undetectable using conventional methods.

Figure A.10: First-Stage: Effect of HEART Implementation on HEART Response Rates



Notes: This figure shows the first-stage regression that estimates the probability of a HEART response, using the implementation of HEART as an instrument. The interaction between time since introduction and eligibility serves as the instrument, with ineligible calls as a control group. Regression controls for [COVARIATES] fixed effects for day, hour, priority, month-year, and beat. The vertical dashed line marks the introduction to HEART. Standard errors are clustered at the address level.