NBER WORKING PAPER SERIES

WHAT IS THE BEST RESPONSE?
EXAMINING THE IMPACT OF POLICE AND THEIR ALTERNATIVES

Bocar A. Ba
Patton Chen
Tony Cheng
Martha C. Eies
Justin E. Holz

Working Paper 34344
http://www.nber.org/papers/w34344

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2025

We are grateful to Amna Akbar, Desmond Ang, Peter Arcidiacono, Pat Bayer, David Berger, Alec Brandon, Kevin Dano, Michael Dinerstein, Andrew Fan, Dalia Ghanem, Nate Hendren, Anjelica Hendricks, John Eric Humphries, Jacob Kaplan, Desire Kedagni, Illenin Kondo, John Macdonald, Matt Masten, Naomi Murakawa, Aurelie Ouss, Michael Pollman, Canice Prendergast, Roman Rivera, Evan Rose, Yotam Shem-Tov, Gabriela Solis, Ben Sprung-Keyser, Megan Stevenson, Atheendar Venkataramani, Marianna Yamamoto, Jonathan Zhang, and the Bellwether Collaborative. Special thanks to Ryan Smith and Keith Chadwell for providing feedback on HEART program. We also thank seminar participants Georgia State, Michigan Ford, Penn LDI, Vanderbilt, Mortality and Criminal Legal System Contact Conference at Cornell, and NBER Summer Institute. Iris Chang and Lily Levine provided outstanding research assistance. The study was approved by the IRB at Duke (IRB 2025-0376). The field experiment was pre-registered in the AEA RCT Registry (AEARCTR-0016060). We gratefully acknowledge financial support from J-PAL North America's Social Policy Research Initiative and Arnold Ventures. This research received support from the Population Dynamics Research Infrastructure Program award to the Duke Population Research Center (P2C HD065563) at Duke University by the Eunice Kennedy Shriver National Institute of Child Health and Human Development. Any errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

What is the Best Response? Examining the Impact of Police and Their Alternatives
Bocar A. Ba, Patton Chen, Tony Cheng, Martha C. Eies, and Justin E. Holz
NBER Working Paper No. 34344
October 2025
JEL No. H10, H4

ABSTRACT

Cities across America are adopting civilian crisis response programs as alternatives to traditional policing, yet causal evidence on their impact and cost-effectiveness is scarce. This paper evaluates Durham, North Carolina's HEART program, which diverts nonviolent 911 calls from police. Using a difference-in-differences design, we find that HEART reduces crime reports, arrests, and response times—primarily through civilian phone and in-person responses, rather than police-civilian co-responses. The program increases future 911 calls, which suggests it fosters public trust. Based on an original contingent valuation survey and applying the marginal value of public funds framework, we conclude that HEART is a fiscally self-sustainable intervention.

Bocar A. Ba
Duke University
Department of Economics
and NBER
bocar.ba@duke.edu

Patton Chen
Duke University
bingjun.chen@duke.edu

Tony Cheng
Duke University
tony.cheng@duke.edu

Martha C. Eies
Duke University
martha.eies@duke.edu

Justin E. Holz
University of Michigan
Ford School of Public Policy
holzj@umich.edu

1 Introduction

Police are the primary—effectively the default—responders for most 911 calls, including calls for nonviolent incidents such as mental health and substance use crises. Faced with concerns about overpolicing, cities have introduced civilian-led crisis response programs to manage some nonviolent emergencies (Akbar, 2020; Kaba et al., 2021; Phelps, 2024), aiming to deescalate crises and reduce unnecessary arrests. However, the impacts of these programs remain largely untested. Do they improve public safety? Which models work best? For whom? Are they cost-effective?

As seen in Figure 1, crisis response programs have increased sharply in recent years. Although early programs, such as CAHOOTS (Crisis Assistance Helping Out On The Streets) in Eugene, Oregon, date back to the 1980s, their recent growth has been spurred by the 2020 protests against police violence (Davis et al., 2024). As cities adopt and scale these programs, it is vital to assess how well they work. Evaluations will help guide policy decisions, such as whether crisis response programs should remain within police departments or be outsourced to community-based agencies (Hart et al., 1997). Investing in police versus civilian response teams requires balancing trade-offs in cost and quality, while weighing benefits such as de-escalation and community trust.[1]

This paper evaluates Durham, North Carolina's Holistic Empathetic Assistance Response Team (HEART) program to address key questions about the effectiveness and value of crisis response programs. Using linked administrative 911 call data from 2021 to 2024 and a difference-in-differences design, we examine the causal effect of HEART on crime reports, arrests, and community engagement. We find that civilian responses reduce arrests and crime reports, particularly for low-level incidents, without increasing escalation risks, and that they disproportionately lower arrests of groups historically overrepresented in the criminal justice system. Furthermore, backup requests spiked in the early months but declined as the program matured, providing further evidence that HEART helps divert cases from the justice system. To assess the program's broader value, we combine evidence from an original contingent valuation survey on residents' willingness to pay with the marginal value of public funds framework, offering a comprehensive assessment of the effectiveness, equity, and cost-effectiveness of HEART.

Fundamental challenges persist in evaluating crisis response programs. 911 calls range from emergencies requiring immediate intervention, such as fires or active shooter incidents, to less dangerous situations, such as homelessness or loitering. Such variation in the type and urgency of 911 calls makes it difficult to assess response effectiveness, and to know whether reduced police deployments reflect genuine safety gains or merely shifts in reporting. Moreover, not all 911 calls require an in-person response; some may be safely resolved remotely by traditional 911 dispatchers or through consultation with mental health specialists (Ba et al., 2024; Fenizia and Kirchmaier,

[1] Similar trade-offs in cost, accountability, and service quality characterize the outsourcing of other criminal justice services, such as prison management (Mukherjee, 2021; Mukherjee and Sanders, 2021; Yegen, 2021).

2024). Evaluating effectiveness must account for different response types, yet most studies do not make this distinction. Beyond evaluation, policymakers must consider costs and benefits to enable meaningful comparisons across jurisdictions and program models.

HEART's design presents a unique opportunity to overcome these challenges. At its rollout, the program integrated multiple crisis response models within a centralized 911 call system within months. These models include civilian phone response by a mental health clinician who responds to qualified calls; in-person civilian response comprising a mental health clinician, a peer support specialist, and an emergency medical technician (EMT); and co-response by a team pairing a civilian mental health clinician with a crisis intervention team–trained (CIT-trained) police officer for higher-risk situations. These teams are dispatched according to standardized nature codes generated by an automated system, which ensures responder assignment is driven by call content rather than dispatcher discretion. Because the automated system was in place before the program's launch and fully integrated with the city's emergency dispatch infrastructure, we can compare the effects of each type of response on public safety outcomes in a unified and internally consistent framework.

We link 911 call records from 2021 to 2024 (obtained via Freedom of Information Act requests) with program and police response data. We identify HEART-qualifying calls by matching the response records to the call nature codes, which enables us to classify calls made before and after the program's launch. Specifically, we identify pre-program calls that would have qualified for a HEART response and post-implementation calls that did receive a HEART response. Our comparison group includes unqualified calls: those involving suspicious activity, unknown problems, or "assist person" requests, which share similar location, temporal, and priority characteristics with qualified calls but fall outside the program's eligibility criteria.

To evaluate the program's causal impacts, we use a nonstaggered difference-in-differences (DiD) design, comparing outcomes from HEART-qualified 911 calls (those that received either a HEART response or a police response post-implementation) to unqualified, police-handled comparison calls. We estimate the average treatment effect on the treated (ATT) for each response type, controlling for pre-treatment characteristics and time-varying factors. Our empirical strategy rests on standard identification assumptions, including conditional parallel trends, no anticipation, and no spillovers (Roth et al., 2023; de Chaisemartin and D'Haultfoeuille, 2025; Ghanem et al., 2023). We also assume consistent call classification and exogeneity of the comparison group. We validate these assumptions with event-study models and explore heterogeneity by HEART response type.

Regarding key public safety outcomes, we focus on the ATTs for both program responses and police responses that qualified for HEART intervention. HEART's rollout led to significant reductions in crime reporting, arrests, and response times. Among calls receiving a HEART response, crime reports declined by 9.2 pp, over 50% relative to baseline, and arrests dropped by 2.8 pp from a 5% baseline. We observe no reductions among police responses, which suggests the effects are driven by the program itself rather than changes in enforcement behavior. Response times decreased for both program and police responses, but the decline is nearly twice as large for HEART responses. Backup requests initially rose by 4.1 pp relative to a 3% baseline, but this increase was concentrated in the early months post-rollout, diminishing over time. The largest reductions correspond to nonviolent and quality-of-life incidents, although violent crime reports and arrests also declined more for calls handled by HEART than for those handled by police. These findings suggest that although civilian crisis teams may require more coordination early in their implementation, they can divert low-risk calls from the criminal legal system, reducing formal enforcement without compromising safety.

Impact varies by call type, reflecting differences in the urgency of the incident and the need for enforcement. Civilian phone and in-person responses led to fewer crime reports and arrests for low-level calls including trespassing, nuisance complaints, and quality-of-life incidents. The effects are smaller for mental health crises and disturbances and are minimal for calls related to violent incidents or domestic violence. In contrast, police responses are associated with an increase in crime reports, especially for ambiguous or high-risk calls.

Beyond immediate outcomes, the program influences how residents engage with emergency services over time. In contrast to declines in 911 use observed after high-profile incidents of police violence (Desmond et al., 2016; Ang et al., 2024), HEART does not deter future calls. Examining trajectories across three escalation metrics—use of force, repeat 911 calls, and violent follow-up events—we find no evidence that civilian crisis teams increase the likelihood of escalation. Use-of-force incidents remain exceedingly rare under both HEART and traditional police responses. Within 31 days, the likelihood of repeat calls is essentially unchanged, indicating that all models resolve initial crises equally effectively. Over longer horizons, addresses served by HEART generate slightly more follow-up calls than those served by police, yet this increased engagement does not translate into higher rates of violence. Indeed, the probability that future calls involve violent offenses declines under all response types. Taken together, the results suggest that civilian teams handle emergencies without escalation and may encourage continued, constructive use of emergency services.[2]

Finally, the program's impact is most pronounced for groups historically overrepresented in arrests (Bor et al., 2018; Ba et al., 2021). Calls receiving a HEART response result in fewer arrests of Black individuals, men, and people aged 25–39, populations with the highest baseline arrest rates. There is no comparable change for calls receiving police responses. These results underscore the potential for civilian crisis response models to reduce contact between the individuals in crisis and the criminal justice system and racial disparities in enforcement, even without explicit targeting.

[2]We also test whether non-police response options increase 911 reliance using repeat usage patterns and call priority as revealed-preference indicators of trust. We find that civilian responses increase the probability of subsequent crisis calls by approximately 25 pp, nearly doubling baseline rates, and are associated with significantly lower call priority, with phone diversions showing the largest decrease. Co-response and police responses show smaller but similar effects, suggesting spillovers.

We next decompose the program's effects by response type. Specifically, we compare fully civilian interventions with those that integrate law enforcement. Civilian phone and in-person responses both reduce crime reports and arrests, though the magnitude of the effect varies. In contrast, co-responses have minimal impact on crime or arrests, with patterns in outcomes resembling those of calls receiving traditional police handling.

These differences are mirrored in the program's cost profile. Civilian phone responses are the least expensive at $32.50 per hour, reducing labor costs by nearly $49 per call. Civilian in-person responses most effectively reduce arrests but are the most expensive at $98 per hour, increasing labor costs by over $30 per call. Co-response sits between the two in cost and effectiveness. Despite the higher price tag for civilian in-person response, Durham has continued to expand this model, signaling a willingness to pay more for responses that divert cases from the criminal legal system.

To evaluate the social value of HEART, we estimate residents' willingness to pay (WTP) and compare it to the program's net fiscal cost. Using a CV survey, a widely used method for valuing nonmarket goods (Cohen, 2000; Esberg and Mummolo, 2018), we elicit how much Durham residents would be willing to pay to sustain the program's operations, capturing perceived benefits such as reduced harm, better long-term outcomes, and greater trust in public safety services. Our design follows best practices from the CV literature (Bishop et al., 2017; Domínguez and Scartascini, 2024) and experimental survey design (Stantcheva, 2023) and includes a list experiment to mitigate social desirability bias (SDB) (Blair and Imai, 2012; Coffman et al., 2017). We find strong public support for HEART: 95% of respondents report a positive WTP for the program, with the mean valuation being $102.91 per year—more than eight times the program's per-resident cost. After adjusting for SDB, we estimate a WTP of $80, six times the direct cost of HEART.

We also estimate the WTP for each of the three types of police alternatives studied in our program. Our findings show that WTP increases with the level of response involvement. We find that the willingness to pay for civilian phone responses is $27.39, the willingness to pay for police-civilian co-responses is $28.27 and the willingness to pay for civilian in-person responses is $30.15. These findings suggest that the public places higher value on more active crisis interventions. The individual estimates can be used in future research to provide credible WTP estimates for police alternative programs that utilize any of these types of programs.

We assess cost-effectiveness using the marginal value of public funds (MVPF) framework (Hendren, 2016; Hendren and Sprung-Keyser, 2020), which compares an intervention's social benefits to its net fiscal costs. The average HEART response costs $1,191 but generates estimated fiscal savings of $2,093 per call, largely through reductions in arrests and crime-related expenditures. These estimates rely on standard valuations of crime costs from Miller et al. (2021), which include both tangible and intangible costs. The resulting net savings of $902 per call implies an MVPF of infinity, which means that the program pays for itself through fiscal externalities. Together with the high

WTP, these findings suggest that HEART is both welfare-enhancing and self-sustaining, delivering substantial economic and social returns.

Literature Review Our study contributes to research on public safety, crisis response, and resource allocation. A dense police presence can reduce crime (Vidal and Kirchmaier, 2018; Chen et al., 2023), but negative interactions between civilians and police may erode trust and deter reporting (Desmond et al., 2016; Ang et al., 2024). Deescalation training for police officers shows some promise (Owens et al., 2018; Roth and Sant'Anna, 2023; Ross and Sloan, 2024), yet how effective police can be as first responders for mental health crises or substance use remains uncertain (Ba et al., 2024). Alternative models such as virtual crisis response (Gates et al., 2024), have shown potential, but evaluations of interventions of this kind in the US are limited. Our study examines whether shifting emergency response from police to civilian crisis teams improves outcomes, building on evidence that specialized responders reduce police deployment while enhancing crisis resolution (Dee and Pyne, 2022; Deza et al., 2023; Davis et al., 2024). [3]

Finally, 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 examines the welfare impacts of victimization (Stevenson and Mayson, 2022), we extend this research by directly measuring public preferences for police alternatives. Using a CV approach with a multiprice list (Anderson et al., 2007; Allcott and Kessler, 2019), we estimate WTP and assess trade-offs between traditional policing and alternatives. Our method captures both use and nonuse values, offering a more comprehensive view of public preferences than other evaluation techniques. Additionally, we evaluate the fiscal externalities of these alternatives, providing a framework for quantifying 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 reallocation of resources from policing to civilian-led crisis response.

Plan The remainder of the paper is organized as follows. Section 2 provides background on emergency response in the US and Durham's HEART program. We also outline our conceptual framework for evaluating police alternatives. Section 3 describes the data and presents key descriptive statistics. Section 4 details the research design and identifying assumptions. Sections 5 and 6 examine the impact of the program and the different response models on public safety, trust, and labor costs. Section 7 evaluates the program's cost-effectiveness using results from our CV survey and the MVPF

[3] We contribute to the literature on public resource allocation and police funding, including research on law enforcement budgets (Bursztyn et al., 2023; Moreno-Medina et al., 2025), and the broader effects of funding distribution on communities (Derenoncourt, 2022; Cox et al., 2025). By assessing the effectiveness and cost-efficiency of nonpolice crisis response, we inform policy discussions on optimizing emergency response and resource allocation.

framework. The final section concludes.

2 Background on Emergency Response

2.1 Crisis Response Programs in the US

The 2020 murder of George Floyd accelerated efforts to rethink public safety (Akbar, 2020; Kaba et al., 2021; Phelps, 2024), spurring a wave of crisis response programs aimed at reducing police involvement in mental health, substance use, and quality-of-life emergencies. These programs vary widely in design, which complicates their evaluation. Some, such as Eugene's CAHOOTS and Minneapolis's Behavioral Crisis Response (BCR), feature dispatches of unarmed civilian teams (Dee and Pyne, 2022; Davis et al., 2024), while others, such as Seattle's Crisis Response Team, employ a co-response model pairing clinicians with officers. The approaches also differ in their delivery: Alternatives in Houston and Philadelphia include remote 911 diversion models, and cities such as Albuquerque and Austin deploy in-person community-based teams. Most rely on 911 screening, though some, such as the program in Rochester, use dedicated crisis lines (e.g., 211). Despite the existence of over 100 such programs nationwide (Community Safety Workgroup, 2025), we still know little about how effective they are, how well they integrate with other emergency services, and which models work best—questions that we examine systematically in the context of the HEART program here.

2.2 Durham's HEART

HEART Responses In 2022, the Durham Community Safety Department launched HEART to provide alternatives to a police-only response for nonviolent mental health crises and quality-of-life–related concerns, with the aim of reducing unnecessary police involvement[4]. The HEART program consists of four units: Crisis Call Diversion, Community Response Teams, Co-Response, and Care Navigation. Before the program was implemented, all emergency 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 immediate phone and in-person interventions, this arm of the program is not included in our analysis. The three key program arms that we study here are described below:

1. Under Crisis Call Diversion, which we refer to as "civilian phone" response, a telecommunicator first answers the call and uses an automated algorithm that relies on standardized inputs

[4]As of summer 2025, HEART responders are required to submit post-response safety assessments to the Department of Community Safety, with data from the public dashboard indicating they reported feeling safe in approximately 99% of incidents.

to determine what nature code should be applied. For qualifying calls, the telecommunicator transfers the caller to a mental health clinician in Durham's 911 call center. The clinician assesses, and responds to behavioral and mental health-related calls for service—deescalating crises, developing safety plans, and connecting individuals to appropriate care without dispatching responders when possible. If an in-person response is needed, Crisis Call Diversion dispatches the appropriate team, be it the Community Response Team (see below), a co-response team, fire/EMS, or police. Similar programs exist in Houston, Charleston, Austin, and Philadelphia.

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

3. In the Co-Response Team, which we refer to as "co-response," a licensed clinician is paired with a 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 that of programs in many other cities, where co-response programs are housed within police or fire departments, in that it partners the Durham Community Safety Department (DCSD) with law enforcement.

Because one of the program teams (Crisis Call Diversion) operates remotely and the other two (Community Response Team, Co-Response Team) feature on-site intervention, the program's launch provides a natural setting to assess how each of the different crisis response models affects public safety outcomes. By examining Durham's experience of how the presence or absence of police, EMTs, and clinical specialists influences incident resolution, we offer insights into the effectiveness of alternative crisis response strategies and their potential for attending to cases shifted away from traditional law enforcement.[5]

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, this program integrates civilian

[5] For more details, see DCSD.

phone responses, civilian in-person responses, and police–civilian co-response under one framework. This structure allows us to draw direct comparisons between in-person and phone-based interventions and between responses that do and do not involve police, providing a more comprehensive assessment of crisis response strategies.

The spatial distribution of the program's responses provides some indication of the program's reach into 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.[6] Our setting ensures that the evidence on the effectiveness of alternative crisis response is not limited to predominantly white populations, offering insight into the equity and scalability of alternative public safety models. HEART responses occur most frequently in these areas, consistent with their having the greatest need for crisis intervention. The different response models are deployed in distinct geographic patterns: Co-response is deployed most densely in high-crime areas, while civilian phone responses are more evenly distributed but remain concentrated in moderate-crime areas. civilian in-person responses are the most dispersed, reaching both high-crime neighborhoods and lower-density residential areas. The overlap between the program response locations and communities historically subject to a disproportionate police presence suggests the program effectively targets high-need areas. The expanding footprint of civilian in-person responses over time may indicate a growing reliance on nonpolice crisis intervention and an extension of alternative response models beyond the most crime-dense neighborhoods.

2.3 Conceptual Framework

The expansion of alternative crisis response programs raises a question for public safety systems: When and how should police be replaced with civilian-led teams in responding to emergency calls? The answer depends not only on the direct effects of different types of responses—on crime, arrests, or escalation—but also on how these models reshape trust in emergency services, resource allocation, and disparities in enforcement. We develop a simple conceptual framework to guide our empirical analysis and interpretation. Theoretically, the effects of the different types of responses are expected to operate through several mechanisms listed below, with observable outcomes indicated in parentheses:

- Deescalation and diversion (↓ crime reports, arrests): Responses by a civilian team may reduce the likelihood that low-risk incidents result in arrest or criminal charges, particularly when

[6] Durham's demographics set it apart from other cities with nonpolice emergency response: Its demographic composition is 34.4% Black, 14.7% Latino and 43.9% white, in contrast to that of cities with comparable programs such as Eugene (1.7% Black, 4.5% Latino, 79% white) and Denver (8.9% Black, 27.9% Latino, 63% white). Statistics are from U.S. Census Bureau's QuickFacts (accessed August 2025), Durham City, North Carolina; Eugene City, Oregon; Denver City, Colorado.

the core issue is behavioral health, not lawbreaking. This implies that we might expect larger effects for calls related to nuisance, trespass, or mental health.

- Escalation risk and mismatch ($\uparrow$ backup requests): Civilian responses may be suboptimal if the teams are dispatched to attend to calls that exceed their capacity, leading to higher demand for backup. This outcome is likelier early in the implementation period or when call severity is underestimated.

- Service connection and long-run trust ($\downarrow$ future arrests, $\uparrow$ calls): Civilian teams not only may resolve immediate crises but also may help connect individuals to services, lowering future contact with the criminal legal system. In turn, the availability of nonpolice alternatives may encourage help seeking in communities that underuse 911 because they distrust police.

- Reductions in disparities ($\downarrow$ arrests among historically overpoliced groups): If civilian responses disproportionately reduce enforcement against groups likelier to be arrested under the status quo, such as Black men, then the program may reduce disparities even without targeting specific demographic groups.

These mechanisms imply distinct, testable predictions across response types and outcomes. A finding of greater reductions in arrests and crime reports for calls receiving a HEART response than for calls receiving a police response, particularly nuisance and mental health calls, would support the diversion mechanism. A greater increase in backup requests for calls receiving a civilian response, especially soon after program rollout, would suggest transitional mismatch and resourcing issues. If use-of-force and follow-up calls for violent incidents fall across all responses, this would support a reduction in escalation. Finally, if arrest reductions are concentrated among historically overpoliced populations, the program may reduce enforcement disparities even absent an explicit focus on equity in its design.

3 Data and Descriptives

3.1 Data Construction

Source The primary data for this analysis come from public record 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 call nature codes, allowing us to identify incidents qualified for a program response. Geographic details about each call record—the district, beat, census tract, and exact address—provide spatial context, while disposition codes indicate how each case was resolved. The program assignment data specify which HEART arm—the civilian phone response, civilian in-person response, or police–civilian co-response—was dispatched to attend to qualified calls. The arrest data include arrestee demographic information, which allows 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 via case numbers such that, for each observation in our sample of calls, we can determine whether a crime was reported and whether an arrest occurred. Each reported crime is classified as violent, property, or nonindex based on the offense type.[7] 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 calls qualified for a program response by linking the HEART response records with the call nature codes generated by Durham's automated 911 system. When a resident calls 911, the call-taker follows a standardized script, and on the basis of the caller's responses to the preset questions, the system automatically assigns the call one of over 1,000 nature codes. These codes determine the recommended responder type through a preprogrammed dispatch algorithm, leaving no discretion to the individual call-taker. Because each call's eligibility for a HEART response is determined by this structured logic, not human judgment, selective assignment is minimized.[8]

We leverage this structure to distinguish calls that would have qualified for HEART before the program's launch in June 2022 (but that necessarily received a police response) and those that received a HEART response post-implementation. The treatment group includes calls classified as having received a civilian phone response, a civilian in-person response, or co-response (a joint response by a clinician and police officer), with each of these reflecting a different severity level. We also analyze calls where the incidents met the criteria for a HEART response but still received a police response post-implementation. This approach allows us to isolate the effects of alternative vs. traditional crisis handling.

Our comparison group includes calls not qualified for a HEART response—such as calls about suspicious activity, unknown problems, or assist-person requests—where initial uncertainty pre-

[7]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.

[8]See point 12 of the FAQ at www.durhamnc.gov.

cluded immediate qualification. To maintain a clean counterfactual, we exclude cases involving violence or clear public safety threats requiring police by design. This strategy yields a credible and policy-relevant design for estimating the effects of HEART responses relative to the outcomes of comparable calls handled by police in both the pre- and post-periods.

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. Any calls for which we lack 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 that qualified for a program response, 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 treatment assignment by identifying cases assigned to each of the program's units (civilian phone response, civilian in-person response, and/or co-response). The final sample is restricted to cases occurring within designated police districts and beats where the program operates. Additionally, census-based demographic indicators for the incident locations, including racial composition, unemployment rates, and median income, are merged to allow a richer contextual analysis.

The final dataset consists of 161,054 calls, including 43,161 unqualified cases and 103,256 program-qualified cases that received a police response and 14,637 program-qualified cases that received a program response.

3.2 Descriptives

This section examines the implementation and utilization of the alternative crisis response models, tracking their evolution, impact on public safety, and variation across neighborhoods by demographics and incident outcome (crime reports, arrests, time to clear the incident, etc.). Using linked administrative data, we document shifts in patterns of call responses and the extent to which these programs have replaced traditional police handling.

Trends Over Time The top panel of Figure 3 shows monthly trends in HEART-qualified and comparison calls from 2021 to 2024. Prior to the program's launch in mid-2022, all HEART-qualified calls—the number of which per month ranged between 1,800 and 2,500—were handled by police. In contrast, the number of comparison calls per month remained stable at approximately 1,000 throughout the period. Following the program's introduction, the total volume of qualified calls increased, driven by a rise in the number of HEART-handled cases, while the number of police responses to qualified calls remained steady at approximately 2,000 per month. By 2024, the HEART teams were managing over 500 calls monthly. These patterns suggest that the program expanded response capacity rather than displacing police activity, as the volume of comparison calls remained unchanged.

The bottom panel of Figure 3 shows a word cloud of the types of calls handled by each response model. Comparison calls are dominated by calls related to suspicious activity and assist-person calls, indicating that the associated incidents are low-acuity, noncrisis police matters. Calls receiving civilian phone responses primarily involve follow-up, mental health, and crisis call diversion cases, corresponding to low-risk, nonurgent situations. The civilian in-person teams handle a broader range of issues, including disturbances, trespassing, and welfare checks, managing moderate-risk crises without police. The co-response teams are most often dispatched to attend to higher-risk incidents such as drug-related calls, disturbances with weapons, and domestic violence, which underscores the more selective use of responses involving police for complex or potentially dangerous situations.

Summary Statistics Table 1 summarizes the characteristics and outcomes of the unqualified cases, police-handled qualified cases, and HEART-handled qualified cases. Column (1) reports overall means, Columns (2)—(4) disaggregate by case type, Column (5) shows mean differences between the program- and police-handled cases, and Column (6) provides p-values for these differences.

Demographic and economic characteristics are broadly similar across the case groups, with small differences in racial composition and median household income. Qualified cases tend to occur in lower-income areas than unqualified cases, but there is little difference on this metric between the HEART- and police-handled cases. Differences in racial composition across the case types are minor in magnitude and not statistically significant.

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

4 Empirical Strategy

4.1 Target Parameters and Identification

Nonstaggered DiD We first formalize our identification strategy to estimate the effect of a HEART response. To assess the program's impact, we compare the outcomes of HEART-qualified calls to those of unqualified (police-handled) calls. We do not use qualified police-handled calls as a control group because officers may adjust their behavior in response to the program's presence. Instead, we treat both calls actually receiving HEART responses (H) and qualified calls that received police responses to (P) as our treated groups, using unqualified calls as a common comparison group.

HEART was implemented in June 2022 in a nonstaggered rollout. Accordingly, following Roth et al. (2023) and de Chaisemartin and D'Haultfoeuille (2025), we apply a standard DiD design. Following Rubin (1974); Robins (1986), for each call i in period t, let $Y_{it}(0)$ denote the potential outcome if the call is unqualified (i.e., never treated), and let $Y_{it}(d)$ denote the potential outcome if the call is treated with a response of type $d \in \{H, P\}$ after implementation. Define $D_i = 1$ if call i receives treatment d and 0 otherwise. As in standard causal inference frameworks (Holland, 1986), only one potential outcome is observed for each unit. The realized outcome is given by $Y_{it} = D_i \cdot Y_{it}(d) + (1 - D_i) \cdot Y_{it}(0)$. We assume SUTVA, meaning that HEART responses do not influence the outcomes for other, untreated calls—for example, they do not alter how police respond in comparison cases.

The estimand of interest is the ATT for treatment d in the post-period:

$$ATT_d = \mathbb{E}[Y_{i,\text{post}}(d) - Y_{i,\text{pre}}(d) \mid D_i = 1] - \mathbb{E}[Y_{i,\text{post}}(0) - Y_{i,\text{pre}}(0) \mid D_i = 1] \tag{1}$$

This estimand captures the average change in outcomes for calls that received a response of type d—either HEART (H) or police (P)—relative to how those same cases would have been resolved had they been handled similarly to unqualified calls in the absence of HEART. In other words, it measures how introducing treatment d for program-qualified calls, instead of defaulting to a traditional police response, affects key public safety outcomes.

Identification relies on two further assumptions: no anticipation and parallel trends. The no-anticipation assumption requires that treatment status in the post-period not affect outcomes in the pre-period—that is, that $Y_{i,\text{pre}}(0) = Y_{i,\text{pre}}(d)$ for all treated units. In other words, outcomes prior to the program's implementation must be unaffected by future exposure to treatment. The parallel trends assumption states that, in the absence of treatment, the outcomes of qualified and unqualified calls would have followed similar trends over time: $\mathbb{E}[Y_{i,\text{post}}(0) - Y_{i,\text{pre}}(0) \mid D_i = 1] = \mathbb{E}[Y_{i,\text{post}}(0) - Y_{i,\text{pre}}(0) \mid D_i = 0]$.

HEART did not exist in the pre-period, meaning that all qualified calls were handled by police.

Thus, we mechanically impose $Y_{i,\text{pre}} = Y_{i,\text{pre}}(P)$ for all units with $H_i = 1$. In other words, for all calls that eventually received a HEART response after the program began, we assume that in the pre-period, those same calls would have been handled by police.

Conditional Parallel Trends Assumption (PT-X) While the standard DiD design relies on the assumption that the outcomes of the treated and control groups would have followed parallel trends in the absence of the treatment, we follow Roth et al. (2023); Ghanem et al. (2023) and relax this assumption to parallel trends *conditional on observed covariates*. Specifically, we assume:

$$\mathbb{E}[Y_{i,\text{post}}(0) - Y_{i,\text{pre}}(0) \mid D_i = 1, X_i] = \mathbb{E}[Y_{i,\text{post}}(0) - Y_{i,\text{pre}}(0) \mid D_i = 0, X_i] \tag{2}$$

where X_i includes baseline characteristics such as time of day, day of week, location, and call priority. Conditioning on these observables accounts for systematic differences in call composition across groups and strengthens the credibility of our identification strategy. We also test for pre-treatment balance to assess potential violations of this assumption.

4.2 Estimation

Estimating the Effect of HEART For our main analysis, we estimate the sample means analogue of the regression specification in Equation 1, where we allow for covariates with the PT-X assumption (Ghanem et al., 2023). We estimate the following models separately for each type of response, $d \in \{P, H\}$, which we compare to the outcomes of the comparison group (unqualified calls):

$$Y_{it} = \alpha + Post_t \times D_i \times \beta + D_i\lambda + \gamma_t + X_i'\delta + \epsilon_{it} \tag{3}$$

where Y_{it} represents crime reports, arrests, response time, or backup requests for call i and time t. The vector X_i includes covariates from Table 1, along with time-of-day, priority, and location fixed effects. Month–year fixed effects γ_t account for time trends, and errors are clustered at the address level. As discussed previously, to account for broader policing trends, we use unqualified cases—for which police remained the sole responders—as control. This ensures the observed changes reflect the program's introduction rather than differences between qualified and unqualified calls. Here, D_i denotes program-qualified calls, while $Post_t$ equals one after the program's implementation and zero otherwise.

Our key parameters, β, measure how the outcomes of qualified calls changed after the introduction of the program under $d \in \{P, H\}$ responses. Identification again relies on the parallel trends assumption that absent $d \in \{P, H\}$, the trends in outcomes for qualified and unqualified calls would have followed a similar trajectory. To validate this assumption, we estimate a dynamic version of equation 3, allowing the treatment effects to vary over time:

$$Y_{it} = \alpha + \sum_t P_t \times D_i \times \beta_t + D_i\lambda + \gamma_t + X_i'\delta + \epsilon_{it} \tag{4}$$

where P_t represents month–year dummies, with May 2022 taken as the reference period. This specification compares the outcomes of qualified and unqualified cases before and after the HEART program's launch so that we can evaluate whether preexisting differences drive the results. If β_t is small and statistically insignificant for the pre-period, this would offer evidence that the parallel trends assumption holds.

Estimating the Effect of Each Program To better understand how HEART shapes public safety, we disaggregate its effects by evaluating each response model separately: civilian phone response ($H1$), civilian in-person response ($H2$), and police–civilian co-response ($H3$). For each program $d \in \{H1, H2, H3\}$, we estimate a separate DiD model, analogous to Equations 3 and 4, comparing the outcomes for calls qualified for a response of type d before and after the launch of HEART to those of comparison calls over the same period. The pre-period outcomes for each d are drawn from calls that would have been qualified for that specific type of response prior to implementation.

Identification rests on several key assumptions stated previously. First, we assume conditional parallel trends: that absent treatment, the outcomes of HEART-qualified and comparison calls would have followed similar trends. Second, we assume no anticipation, meaning that HEART's rollout did not affect pre-period behavior. Third, SUTVA rules out spillovers, implying, for instance, that HEART responses do not influence police behavior on unrelated calls. Fourth, we assume consistent call coding over time, that is, that HEART did not change how incidents were classified. Finally, we assume exogeneity of the comparison group: that comparison calls are not indirectly influenced by the presence of HEART and, importantly, are never routed to HEART teams in practice.

This identification strategy is supported by the operational structure of Durham's 911 dispatch system. As noted earlier, call classification is determined by an automated algorithm that relies on standardized inputs, removing discretion from call-takers. The HEART response types were fully integrated into the system prior to the program's launch, ensuring that eligibility for a program response is rule-based rather than driven by dispatcher judgment. This design helps rule out compositional changes in call classification following the program's implementation.

5 Police vs. HEART Responses

5.1 Impact of HEART on Incident Outcomes

5.1.1 Impact on Crime, Arrests, Requests for Backup and Response Times

Our main results, shown in Figure 4, reflect the impact of the HEART program on key incident outcomes estimated with our DiD design. The program led to notable shifts in crime reporting, arrests, response time, and calls for backup relative to the outcomes of unqualified calls, with the effects evolving over time. The pretrends appear flat across all outcomes, supporting the parallel trends assumption for a causal interpretation.

There are no meaningful changes in the outcomes of police-handled cases post-implementation, which suggests that the observed effects stem from the program itself rather than changes in officer behavior. Prior to the program, 16% of qualified calls gave rise to the issuance of a crime report. This share remained unchanged for HEART-qualified calls that received a police response post-program (0.003, $SE = 0.004$) but fell by 9.2 pp ($SE = 0.005$), or 57.5%, for calls that received a HEART response. Arrests followed a similar pattern: no change for HEART-qualified police responses (-0.005, $SE = 0.002$) and a 2.8 pp decline ($SE = 0.002$) for HEART responses, from a 5% baseline—indicating fewer formal enforcement actions rather than substitution.

Response times also declined: From a 50.02-minute baseline, response times for HEART-qualified calls handled by police fell by 3.03 minutes ($SE = 0.769$), and response times for calls attended to by HEART fell by 5.89 minutes ($SE = 1.03$), with the effects stabilizing over time. In contrast, backup requests rose for calls handled by HEART. While their pre-program rate was 3%, this slightly declined for qualified calls taken by police (-0.004, $SE = 0.001$) but rose by 4.1 pp ($SE = 0.003$)—more than doubling the baseline—for calls with HEART responses. This spike was most acute in the program's early months and diminished as it matured.

5.1.2 Impact by Type of Crime

Next, in Table 2, we break down the overall changes in crime reporting and arrests after the program's introduction and check the trends in Appendix Figures A.1 and A.2.

Crime reports and arrests fell after the program's implementation, with the largest effects for violent and other crimes. Before the program existed, 8% of qualified cases gave rise to a violent crime report; after the program was launched, this figure declined by 0.8 pp for police responses to qualified calls and by 3.9 pp for calls with HEART responses, where $p < 0.01$ for both coefficients. Arrests for violent crime, with a 3% baseline rate, fell by 1.3 pp after the program's launch for qualified calls with police responses and 1.5 pp for calls with HEART responses. Property crime reports dropped by 0.8 pp from their pre-program baseline of 1% with HEART responses and by

0.4 pp under police responses. Other crime reports, covering 9% of cases, declined by 5.9 pp for HEART responses, with little change for police responses. Arrests followed similar trends, with the program driving larger reductions, particularly for nonindex crimes. Except for the estimates for arrests for property crime, the differences between the HEART and police outcomes are statistically significant at the 1% level.

Overall, HEART responses systematically and largely decreased formal enforcement, particularly for lower-level and quality-of-life incidents. These patterns suggest that civilian-led response alternatives may reduce the likelihood that such cases end up in the criminal legal system, even in the absence of changes in officer behavior.

5.1.3 Impact by Nature of the Call

The program'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 the program diverts cases from the criminal justice system, increases formal enforcement, or introduces trade-offs in safety and efficiency. Figure 5 presents estimates by call type: mental health, trespass or nuisance, disturbance, violent or domestic violence incidents, and other types.[9]

HEART responses consistently reduce crime reports, with the largest declines observable 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 reports 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. In contrast, we observe little change from baseline in arrests after police responses, which suggests that enforcement patterns remain largely unchanged.

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

[9] 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 corresponding estimates more complex.

5.1.4 Do HEART Responses Lead to Escalation?

Crisis calls, if mishandled, can generate significant externalities, from community distrust to legal risks (Bor et al., 2018; Ang, 2020; Ang et al., 2024). Next, we examine whether HEART responses affect the likelihood of escalation, including future 911 calls and use-of-force incidents. If the program discourages follow-up calls, it could signal mistrust in public safety or unresolved needs. Conversely, sustained or increased engagement may suggest that HEART builds trust without increasing punitive outcomes.

Table 3 assesses whether HEART responses reduce the risk of escalation more than police responses. We evaluate three outcomes that we take as indicators of escalation: use of force, repeat 911 calls within 31 days, and violent follow-up incidents. For all outcomes, HEART responses are associated with reductions in use of force, but these reductions are not statistically significant. Similarly, there is no meaningful change in the likelihood of a repeat 911 call within 31 days. Overall, while the point estimates suggest potential benefits, the effects are not statistically distinguishable from zero. Over the longer horizon (more than 31 days), HEART responses increase the likelihood of new calls by 4.3 pp ($SE = 0.0077$), slightly more than the 3.2 pp increase from police responses ($SE = 0.0060$). For future calls involving violent crime, both types of responses lead to declines (HEART: –0.62 pp, $SE = 0.0021$; Police: –0.4 pp, $SE = 0.0018$). These results suggest that relative to a police response, HEART responses do not increase the risk of escalation and may foster willingness to continue to engage with emergency services without increasing the use of force or violent outcomes. Appendix Figures A.3 and A.4 present event-study estimates that show no significant differences in pretreatment trends, supporting the validity of the DiD design.

5.2 Arrestee Profiles and HEART

Table 4 examines the impact of the program 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 the program changed these dynamics is key. The event-study estimates in Appendix Figures A.5 and A.6 confirm no differential pretrends, supporting the validity of our DiD approach.

The top panel shows arrest rates before the launch of the program 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 the program significantly reduced arrests among Black individuals (-2.4 pp, $p < 0.01$), men (-2.2 pp, $p < 0.01$), and people aged 25–39. The effects for other groups are smaller and not statistically significant, suggesting that the program's largest impact is on populations historically overrepresented in arrests. These declines indicate that the program 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 contact with law enforcement, especially for groups with higher baseline arrest rates. The demographic composition of the reference group, individuals processed by police, confirms that Black individuals and men remain most affected, which reinforces how alternative crisis response models can reduce law enforcement contact among historically heavily policed populations (Ba et al., 2021; 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 of our program impact estimates 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% confidence intervals (CIs). Solid lines indicate HEART responses, and dashed lines represent police responses.

Across all outcomes—crime reports, arrests, response time, and backup requests—the bounds confirm the robustness of our DiD estimates. The reduction in crime reports remains significant under conservative assumptions. Arrests also decline, with the bounds ruling out large positive effects. The program reduces response time but increases backup requests, though we see no evidence of overestimation. For police-handled cases, the estimates are noisier, with the confidence intervals widening as the restrictions increase, such that the estimates for crime reports and arrests are less precise.

Overall, the program effects remain negative and meaningful under reasonable assumptions; these results reinforce our confidence that it reduces formal enforcement without compromising safety. The police response estimates are more sensitive to violations of parallel trends. However, the broader pattern confirms that the program consistently lowers crime reports, arrests, and response times, while the effects for police are less clear.

Event Study and Hypothesized Trends Figure A.8 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 pretrends in crime reports, arrests, or backup requests, confirming that the observed post-treatment effects are not driven by preexisting differences. The dashed lines represent the event-study coefficients we would expect if deviations were present but undetectable. Notably, the program's impact on arrests and response times remains statistically significant even under conservative assumptions about violations of parallel trends.

5.3.2 Impact of HEART According to TSLS

As an alternative to our DiD design, we estimate the effect of receiving a HEART response in a two-stage least squares (TSLS) approach, instrumenting treatment with the interaction of program rollout and call qualification for a HEART response ($Post_t \times Qualify_i$). This captures the effect of HEART for calls that we classify as treated on the basis of the program's introduction and targeting criteria. Appendix A provides full details on the estimation strategy, including the first- and second-stage specifications and tests of instrument strength and balance.

Table A.5 presents the impact of the HEART program on a range of outcomes under each model, whether civilian phone, civilian in-person, or co-response. To assess robustness, we compare our estimates from the main DiD specification to those from the TSLS model, which recovers different estimands. Across most outcomes, the TSLS estimates are directionally consistent with the main results but generally smaller in magnitude and less precise. For example, the main analysis finds that HEART reduced the total response time by 5.4 minutes ($SE = 1.03$), while the TSLS estimate is -21.6 minutes ($SE = 5.12$). For overall crime reporting, the main estimate is a 9.4 pp reduction ($SE = 0.005$), but it is 7.9 pp under TSLS ($SE = 0.024$). For arrests, the TSLS estimates are slightly larger: The effect on arrests for violent crime is -1.3 pp in the main model vs. -3.5 pp under TSLS. Though the TSLS estimates are less precise, the consistency in sign and relative magnitude across outcomes supports the robustness of the main findings.

6 Impact of Each Program

6.1 Civilian Phone, Civilian In-Person, and Police-Involved Responses

To better understand how the program affects public safety, we evaluate its impact across the different response models: civilian phone, civilian in-person, and co-response by a police officer and a civilian mental health professional. This distinction allows us to compare responses that fully remove police from crisis intervention with those in which law enforcement remains involved. We employ a DiD approach, comparing the outcomes of cases receiving each type of response to the outcomes of unqualified cases to estimate the relative impact on crime reports, arrests, response times, and backup needs using Equations 3 and 4.

Figure 6 shows that both civilian phone and civilian in-person responses reduce crime reports and arrests, albeit by differing magnitudes. Civilian phone response leads to a 3.5 pp reduction in crime reports relative to their frequency for unqualified calls ($SE = 0.010$), while civilian in-person response reduces crime reports by 7.3 pp ($SE = 0.004$). Civilian in-person responses reduce arrests (-1.9 pp, $SE = 0.002$), while civilian phone responses increase arrests relative to their rate among unqualified cases (1.3 pp, $SE = 0.003$). This finding of an increase in arrests for civilian phone responses may be linked to the initial challenges in fully resolving incidents remotely, as reflected in the surge in backup requests for civilian phone responses (7.6 pp, $SE = 0.018$) in the early months post-rollout. Many civilian phone responses required additional intervention by in-person teams, leading to police involvement and potential arrests. Over time, backup requests for civilian phone responses declined, suggesting better call triage and more confidence in remote handling of crises, which may have mitigated the need for escalation.

Response time patterns further differentiate the performance of the two civilian-only models. Civilian phone response significantly reduces the total response time relative to that of unqualified calls (-25.89 minutes, $SE = 2.44$), while civilian in-person response increases the response time by 5.67 minutes ($SE = 1.02$) and co-response extends the response duration by 4.70 minutes ($SE = 1.55$). These findings reflect the efficiency of remote interventions and the more challenging, time-intensive character of the interventions by in-person crisis teams. However, the early increase in civilian phone response backup requests, coupled with the smaller reduction in crime reports and increase in arrests for this type of response, 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 civilian in-person response, despite the longer response times, suggests that in-person teams are more effective in deescalating incidents on-site without law enforcement intervention.

Overall, the program's crime prevention and arrest reduction effects are concentrated in civilian-only responses, with civilian in-person response yielding the largest declines relative to the outcomes of comparison calls. The stabilization of backup requests from civilian phone responders over time suggests that the initial challenges in remote crisis management were ameliorated, but the smaller impact on arrests for responses of this kind indicates the limits of phone-based interventions in fully deescalating certain situations. Meanwhile, the minimal effects of co-response on crime reports and arrests relative to the outcomes of unqualified calls suggest that when police are present, response patterns largely mirror those under traditional enforcement. These results highlight the key distinction between phone 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 Labor Costs

The expansion of the program 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 7 highlights these differences, breaking down the hourly costs of each type of response, the scale of calls handled, and the overall impact on labor cost. The top-left panel shows that police responses cost $67.53 per hour while the cost of the program's models varies significantly: Civilian phone response is the least expensive at \$32.50, civilian in-person response the most expensive at \$98.09, and co-response in between at \$73.90. The top-right panel reveals the distribution of calls by the type of response received—police responses remain dominant, with police handling 103,256 calls, while 8,909 receive an civilian in-person response, 2,857 a civilian phone response, and 2,871 co-response.

The bottom panel quantifies the program'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 that civilian phone response reduces costs by \$48.90 per call while civilian in-person response and 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.10 presents the event-study versions of these estimates, confirming that the pretrends for the period before the program's implementation are similar.

The cost dynamics shift when we look at total expenditures. Civilian phone response generates the largest per-call savings because of the lower wages—since only one employee handles each call, in contrast to the multiperson teams used in the civilian in-person and co-response models—and the smaller volume of calls receiving responses of this type, but its overall budget impact remains limited. Civilian in-person 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 civilian in-person teams suggests a shift toward civilian-led crisis response despite its higher price tag. While expanding civilian in-person response raises costs in the immediate term, savings from fewer arrests, lower incarceration rates, and reduced crime could outweigh these expenses. If the program 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.

6.3 Does HEART Improve Trust in 911?

Researchers often infer trust in the police by analyzing public reactions to high-profile incidents of police violence (Desmond et al., 2016; Rivera and Ba, 2023; Ang et al., 2024) or through self-reported survey data (Tyler, 2001; Pickett et al., 2022; Cheng and Liu, 2025). The introduction of HEART offers a unique opportunity to examine whether alternative crisis response models influence community engagement with 911. A shift toward lower-priority 911 calls over time may indicate growing trust in emergency services, as community members feel more comfortable reaching out for help with less urgent situations—not just serious emergencies.

Follow-Up Calls at the Same Address Panel A of Figure 8 plots monthly DiD estimates for two outcomes measured at the same address following a 911 incident: (i) whether a new HEART-qualifying call occurs in the future[10] and (ii) whether a new call with the identical nature code arrives in the same window.

Civilian responses show the strongest effect on the first outcome: civilian phone and in-person raise the probability of a new qualifying call by at least 22 pp ($p < 0.01$), almost doubling the pre-program baseline of 0.26. Police-involved responses also increase the probability of new qualifying calls by at least 16 pp.

However, the pattern is different when looking at follow-up calls with the same nature code. The civilian phone response increases this probability by about 9 pp, indicating that residents feel assured they will be helped if they call 911 once they have experienced a remote, unarmed intervention. By contrast, the civilian in-person team produces a small but statistically significant decline by about 1 pp. Co-response shows no discernible effect and police-only cases register a negative shift of 1.9 pp. In short, civilian phone responses appear to encourage repeat engagement with emergency services for similar crises, while its in-person teams are associated with a small reduction in such calls. Police-centered approaches show either no impact or a suppressive effect on repeat calling.

Shifts in Call Urgency A priority code is assigned at the time a 911 call is logged. In Durham's Computer-Aided-Dispatch system, higher numeric values denote lower urgency. For ease of interpretation, we reverse-coded the metric so that it now runs from 0 (least urgent) to 5 (most urgent). Viewed through the lens of public behavior, a downward shift in the priority of future 911 calls from the same address can be interpreted to mean that residents are willing to dial 911 for problems that feel less pressing, one benchmark of increased trust in the system.

Panel B of Figure 8 uses the same DiD framework but now treats the priority score itself as the outcome. The resulting estimates are best viewed as documenting the relationship between HEART and call urgency. Calls handled by the civilian phone team have priority scores more than one point lower than similar unqualified calls, whereas those routed to civilian in-person teams are about half a point lower. Co-response and police-only HEART-qualifying incidents also register negative gaps, though the magnitudes are modest. In practical terms, HEART's unarmed civilian channels appear to bring a new subset of lower-stakes crises into the formal emergency system, while police-involved options remain focused on higher-priority incidents. Taken together, these descriptive results support the view that expanding civilian alternatives normalizes 911 use for everyday problems and, in doing so, may build public confidence in emergency services.

[10] New call at the same address after 24 hours relative to the focal call.

6.4 Robustness

6.4.1 DiD Assumptions

Bounds on the Relative Magnitudes for DiD According to Figure A.11, the core findings remain consistent across a range of sensitivity assumptions for all outcomes—crime reports, arrests, total response time, and backup requests. Calls receiving civilian phone and civilian in-person responses consistently have fewer crime reports, with bounds that exclude null effects across all magnitude restrictions. In contrast, cases with police and co-responses show no significant differences from unqualified cases in the issuance of crime reports. In terms of response time, civilian phone response yields clear reductions, while the estimates for other models, including police, are imprecise and overlap with zero. Backup requests significantly increase for civilian phone response, while we observe no meaningful change for the other response types.

These results suggest that civilian models—particularly phone response—generate consistent reductions in enforcement outcomes without evidence of overreaction or compromised safety. However, the increase in backup requests for civilian phone responses indicates that additional support is sometimes needed. In contrast, cases with a police-only response or co-response show no robust effects across key outcomes, and the estimates are more sensitive to deviations from parallel trends.

Event Study and Hypothesized Trends Figures A.12 and A.13 present event-study estimates and sensitivity checks for each HEART program to assess potential violations of the parallel trends assumption, using the approach previously introduced from Roth (2022).

Across all response types—civilian phone, civilian in-person, and co-response—we find no significant evidence of differential pretrends in crime reports, arrests, or backup requests. For total response time, the estimates show greater variability, particularly for co-response, but the post-treatment effect for civilian phone response remains robust. The dashed lines represent the path of the event-study coefficients we would expect if undetectable deviations from parallel trends were present. Overall, these results support the validity of the parallel trends assumption.

6.4.2 Impact of Each Program According to TSLS

We also estimate the effect of receiving each type of HEART response using a just-identified TSLS model. We instrument assignment to HEART with the interaction between the timing of program rollout and indicators that a call qualifies for a response of a specific type. This captures the effect of a call being assigned a HEART responder based on eligibility criteria after program rollout. Appendix A provides full details on the identification strategy, first- and second-stage specifications, and instrument strength tests.

Table A.7 presents TSLS estimates for four key call-level outcomes. Across most of the outcomes, the results are qualitatively consistent with the DiD estimates but generally of larger magnitude and lower precision. For example, relative to comparison cases, cases assigned a civilian phone responder have a significantly lower likelihood of crime report issuance by 18.5 pp ($SE = 0.046$) and a total on-scene time that is nearly 59 minutes shorter ($SE = 9.13$). Civilian in-person responses lead to a statistically significant decline in crime reporting and an increase in time spent on scene, consistent with their more service-oriented role. The effects of co-response are more variable: The estimate for total time is large and negative (-126.2, $SE = 27.7$), while the estimates for the other outcomes are imprecise and not statistically significant. Overall, while the TSLS estimates are noisier, the signs and relative magnitudes align with the main DiD findings.

7 Welfare Implications

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

7.1 Government Costs

To benchmark the program's value to residents, we first establish the fiscal operating costs, including personnel, training, and administrative costs. Over the analyzed period from 2021 to 2024, the total direct cost of the program amounted to \$17,433,076, equivalent to \$1,191 for each of the 14,637 HEART responses in our sample. This budget reflects a steady increase in funding, which rose 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 of the program 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 their WTP and comparison to the program's net fiscal cost under the MVPF framework.

7.2 Measuring Public WTP with a CV Survey

Although alternative crisis response programs have expanded rapidly, their perceived value remains unquantified. We address this gap by estimating Durham residents' WTP—the income reduction that would leave them indifferent to the program's presence—as a measure of its social value.

WTP reflects the residents' valuation of four key benefits from the program: (i) reduced harm to individuals in crisis (e.g., fewer arrests and less trauma), (ii) improved long-term outcomes via fewer criminal records, (iii) reduced burdens on jails and courts, and (iv) greater public trust and community well-being.

We estimate WTP using a preregistered contingent valuation (CV) survey, a widely adopted approach for valuing nonmarket goods. In contrast to traditional crime control valuations that exclude offender utility from social welfare calculations (Trumbull, 1990; Cohen, 2000; Chalfin, 2015), our framework accounts for HEART's broader social benefits. CV enables an ex ante valuation of both tangible and intangible program effects without relying on exogenous variation, and respondents are informed about the program before reporting their valuations (Cohen, 2000; Esberg and Mummolo, 2018). To mitigate concerns about hypothetical bias, budget neglect, and sample validity, we follow best practices from both the CV (Bishop et al., 2017; Domínguez and Scartascini, 2024) and survey experiment literatures (Stantcheva, 2023).[11]

Sample and Logistics The target population for this study is residents of Durham, North Carolina, who contribute to the tax base that funds the HEART program. We recruited online respondents through YouGov, following online sampling methods similar to those used by Kuziemko et al. (2015). To ensure data quality, we incorporated several validation steps throughout the survey, including multiple attention checks, a comprehension check confirming understanding of HEART services, and a bot-detection question that prompted respondents to copy and paste copyrighted content into a text box. Respondents also submitted demographic information (see Table A.1) and affirmed that they were not AI.

Survey Structure To enhance consequentiality, the survey told participants that their responses would be shared with Durham mayor Leonardo Williams, policymakers and local law enforcement to inform future decisions (Bishop et al., 2017). Respondents received a description of the HEART program and were required to correctly answer four comprehension questions to proceed.

After passing the comprehension check, the respondents were randomized into three initial WTP brackets within an iterative multiple price list (iMPL) design. In this task, the survey asked subjects to make a series of four binary voting decisions about paying a tax to continue HEART for another year, with values ranging from $3 to over $108. We used nonfocal numbers to prevent subjects from bunching at round numbers (Bishop et al., 2017). Following (Domínguez and Scartascini, 2024), we randomly assigned the initial value between subjects that each respondent considered to be $21, $46 or $81 and emphasized the opportunity cost of voting for the policy in terms of

[11] We did not conduct an information provision experiment because our focus is on the willingness to pay (WTP) among community members served by HEART, not the effect of information about HEART on the demand for HEART.

foregone consumption.[12] Each response adjusted the next amount offered up or down, placing participants in one of 12 WTP ranges. After respondents' WTP was determined, the survey asked them to allocate their stated WTP across the three HEART programs.

Next, we used a list experiment to assess the truthfulness of the respondents' self-reports of their WTP and mitigate concerns about social desirability bias (SDB) (Blair and Imai, 2012; Coffman et al., 2017; Cantoni et al., 2019). In this section, we randomly assigned respondents between a control group that received four neutral statements and a treatment group that received the same four statements plus a sensitive fifth item: that the subject would be willing to pay their stated WTP to fund the HEART program.[13] Participants indicated how many of the statements applied to them but not which ones, allowing them to express support for HEART without directly revealing it. The difference in mean responses between the groups provides an estimate of the share of respondents who truthfully support the program at that price point. The full survey can be found here.

Construction of Hypothetical WTP We construct the WTP from the iMPL responses using the methods from Allcott and Kessler (2019) and Holz et al. (2024). Each subject's answers to the four pairwise decisions map into one of twelve WTP intervals: $(-\infty, 0], (0, 3]$, $(3, 12]$, $(12, 21]$, $(21, 34]$, $(34, 46]$, $(46, 56]$, $(56, 69]$, $(69, 81]$, $(81, 94]$ $(94, 108]$, and $(108, \infty)$. Figure A.14 presents an example structure of the iMPL and describes how a respondent's choices lead to one of these twelve intervals.

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 \$7.50 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.[14] This gives \$-10.50 as the conditional mean on $(-\infty, 3)$ and \$127.11 as the conditional mean on $(108, \infty)$.

Correcting for Hypothetical Bias One limitation of contingent valuation (CV) studies is that stated willingness to pay (WTP) may overstate true WTP, as some respondents may feel social pressure to indicate support for policies, even if they would not do so in an actual decision-making context.(Diamond and Hausman, 1994; List, 2025; Bursztyn et al., 2025). To adjust for this, we model respondents as belonging to one of two types: those who tell the truth and those who overstate their WTP. Let $\lambda \in [0, 1]$ denote the share of respondents who truthfully report their WTP and $1 - \lambda$ the share of respondents who overstate their WTP. Then, using the law of total

[12]Table A.1 shows balance on baseline characteristics.

[13]We present subjects with the minimum value of their elicited WTP range unless they stated a WTP of \$0, in which case we present them with the WTP of \$0. The four neutral statements were: (1) "Over the week I usually read at least one newspaper or magazine" (2) "I want to see Durham as a city with a high standard of living" (3) "I know the name of the governor of North Carolina" and (4) "Durham has fairly high quality restaurants. We presented all five statements in a random order. Table A.2 shows balance on baseline characteristics and MPL treatment assignment.

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

expectations we can express the subjects' true willingness to pay as

$$\mathbb{E}[\text{True WTP}] = \lambda \cdot \mathbb{E}[\text{True WTP}|\text{Truth-teller}] + (1-\lambda) \cdot \mathbb{E}[\text{True WTP}|\text{Overstater}]. \tag{5}$$

Because truth-tellers honestly report their WTP, $\mathbb{E}[\text{True WTP}|\text{Truth-teller}] = \mathbb{E}[\text{Stated WTP}|\text{Truth-teller}]$. Next, assume that overstaters state WTPs drawn randomly from the same distribution as truth-tellers, but have a true WTP of 0. Then, we can rewrite Equation 5 as only a function of the average stated WTP and the portion of subjects who honestly report their true WTP.

$$\mathbb{E}[\text{True WTP}] = \lambda \cdot \mathbb{E}[\text{Stated WTP}]. \tag{6}$$

We estimate λ using the list experiment by comparing the average number of statements that apply to the respondents in the treatment group (S_t) to the average number of statements that apply to the respondents in the control group (S_c), $\lambda = \mathbb{E}[S_t] - \mathbb{E}[S_c]$. Additionally, we can estimate λ separately for each WTP range to recover an SDB-adjusted demand curve for HEART.

WTP Estimates The top panel in Figure 9 presents results from the list experiment described above. In the control group, on average, respondents stated that 2.86 of the statements applied to them. In the treatment group, on average, respondents stated that 3.64 statements applied to them. The estimated difference between the treatment and control groups implies that 78% ($p < 0.01$) of the respondents truthfully reported their WTP.

The bottom-left panel of Figure 9 presents our estimates of the respondents' WTP for the HEART program. The black line presents the demand curve estimated using the WTP values elicited from the iMPL. The dotted grey line presents the demand curve adjusted for SDB using the list experiment. To construct the SDB-adjusted demand curve, we calculate the portion of overstaters in each range and set those WTP values to $0.

This figure shows that 79-95% of respondents have a positive WTP for HEART and 74-93% are willing to contribute more than the program's expected per-resident cost in 2024. The average WTP is $102.91 per year ($0.28 per day) or $80 per year ($0.21 per day) after adjusting for SDB. These estimates correspond to roughly six to eight times the direct costs of the program. This strong demand indicates broad and substantial support for HEART, even at meaningful personal cost.

The bottom-right panel breaks down WTP by response type. The respondents expressed the highest WTP for civilian in-person response ($30.15), followed by co-response ($28.27) and civilian phone response ($27.39). Adjusted for SDB, these estimates are $29.37, $22.05, and $21.36, respectively. Demand for civilian in-person response is 8-12% higher than for the other two services

($p < 0.05$) These differences align with the intensity of the response models: Subjects place greater value on more direct, in-person interventions. Notably, WTP rises with the costliness of the response type, indicating that preferences are consistent with the underlying program intensity and resource use.

Together, these estimates point to a public valuation of HEART that is both high and credible; indeed, we find robust support across valuation methods and program types. WTP for the program consistently exceeds its costs, even before we account for its fiscal externalities.

7.3 Measuring Cost-Effectiveness with MVPF

Overview of MVPF The MVPF quantifies a policy's cost-effectiveness by comparing the benefits individuals receive from the program to its net fiscal cost. This indicator captures the net social benefit the program generates per additional dollar of government spending. Thus, the MVPF framework offers a rigorous way for us to evaluate HEART's overall welfare impact. The MVPF is defined as:

$$MVPF = \frac{\text{Beneficiaries' Willingness to Pay}}{\text{Direct Cost of HEART} - \text{Fiscal Externality}} \tag{7}$$

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

The denominator consists of two components. The first is the direct cost of operating the program, 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 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.

If the MVPF exceeds one, the program's benefits exceed its costs. If a positive fiscal externality exceeds the direct cost of the program and the WTP is positive, then we say that the $MVPF = \infty$. This would suggest, in other words, that the fiscal savings from the program fully offset its direct costs, meaning that the program effectively funds itself through reductions in criminal justice expenditures.

Fiscal Externality Motivated by our consistent finding that the HEART program 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 that lead to changes in public safety costs.

To calculate the fiscal externality of the program, we account for government costs and savings using the estimates from Miller et al. (2021) on the economic costs of crime in the US as of 2017, which cover both reported and unreported incidents. These costs span tangible components (e.g., medical care, lost productivity, property damage, public services) and intangible losses, such as diminished quality of life. We report these estimates in Appendix Table A.8. For each call i with response $r \in$ Police, Civilian Phone, Civilian In-Person, Co-Response and crime type k, we compute the total costs as follows:[15]

$$\begin{aligned}\text{Total Costs}_i^{r,k} = &\ \text{Crime}_i^k \times (\text{Medical}_i^k + \text{Mental Health}_i^k + \text{Productivity Loss}_i^k \\ &+ \text{Property Loss}_i^k + \text{Public Services}_i^k + \\ &+ \text{Adjudication \& Sanctioning}_i^k + \text{Perpetrator Work Loss}_i^k)\end{aligned} \tag{8}$$

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

To estimate the average change in total costs following the introduction of the program, we employ the DiD model from Equation 3, where the dependent variable is $\text{Total Costs}_i^{r,k}$. This approach recovers $\hat{\Delta}_{HEART}^{\text{Costs}}$, capturing the cost savings or increase arising from HEART responses to qualified calls, i.e., the fiscal externality.

MVPF Estimates Figure 10 illustrates the impact of the program's introduction on the cost per call, capturing both the program's direct costs and its broader fiscal externalities. Our analysis reveals an overall cost reduction, with the largest components of the savings being quality-of-life improvements, reductions in tangible costs, and smaller work losses on the part of perpetrators, leading to a total crime cost savings of approximately $2,093. In turn, the direct operational cost of a HEART response is $1,191, a figure more than offset by the fiscal savings, resulting in a net government cost impact of -$902. This suggests that the program delivers substantial fiscal benefits by reducing crime-related expenses, including victimization, investigation, and incarceration costs. Furthermore, given the per-response WTP of $5,872, the program generates positive net benefits, implying an MVPF of infinity (∞), as the fiscal savings fully cover the 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 relative to the outcomes of calls not qualified for a program response. These effects are primarily driven by the program's fully civilian teams.

[15] For brevity, we omit the subscripts for the address c and time t.

Police–clinician co-response, on the other hand, has minimal impact. The alternative forms of emergency response offered by the program lead to a shift in the handling of behavioral health crises and quality-of-life incidents away from police without increasing crime. Although civilian phone responders initially issued more requests for backup when the program was first rolled out, these needs diminished over time. Using a novel contingent valuation survey, we find strong public support for this program: Willingness to pay for HEART exceeds its net cost. Applying the marginal value of public funds framework, we find that the fiscal savings from HEART exceed its direct costs, which indicates that it effectively funds itself through externalities.

These findings contribute to the growing policy discussion on the role of police in public safety and the need for non–law enforcement alternatives. However, key questions remain. First, 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 its different response models, but many cities operate with fragmented dispatch systems. Future research should explore whether cities with different crisis response structures enjoy similar gains and how alternative programs affect disparities in policing outcomes at scale.

Our study provides a replicable framework for evaluating nonpolice emergency response alternatives, combining administrative data, natural experiments, and public valuation methods to assess both effectiveness and fiscal sustainability. As cities continue to invest in such emergency response models, rigorous evaluation is critical to ensuring that these programs deliver on their promise of enhancing both safety and equity in 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. (2020, 09). The Effects of Police Violence on Inner-City Students. *The Quarterly Journal of Economics*.

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.

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.

Blair, G. and K. Imai (2012). Statistical analysis of list experiments. *Political Analysis 20*(1), 47–77.

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.

Bursztyn, L., I. K. Haaland, N. Röver, and C. Roth (2025). The social desirability atlas. Technical report, National Bureau of Economic Research.

Cantoni, D., D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2019, 01). Protests as strategic games: Experimental evidence from hong kong's antiauthoritarian movement*. *The Quarterly Journal of Economics 134*(2), 1021–1077.

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

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*.

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

Coffman, K. B., L. C. Coffman, and K. M. M. Ericson (2017). The size of the lgbt population and the magnitude of antigay sentiment are substantially underestimated. *Management Science 63*(10), 3168–3186.

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*.

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

de Chaisemartin, C. and X. D'Haultfoeuille (2025). Credible answers to hard questions: Differences-in-differences for natural experiments.

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.

Diamond, P. A. and J. A. Hausman (1994). Contingent valuation: is some number better than no number? *Journal of economic perspectives 8*(4), 45–64.

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.

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

Fenizia, A. and T. Kirchmaier (2024, Sep). Not incentivized yet efficient: Working from home in the public sector. CEP Discussion Papers dp2036, Centre for Economic Performance, LSE.

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

Ghanem, D., D. Kédagni, and I. Mourifié (2023). Evaluating the impact of regulatory policies on social welfare in difference-in-difference settings.

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.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association 81*(396), 945–960.

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.

List, J. A. (2025). The experimentalist looks within: Toward an understanding of within-subject experimental designs. Technical report, National Bureau of Economic Research.

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.

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.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling 7*(9), 1393–1512.

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.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*, 688–701.

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.

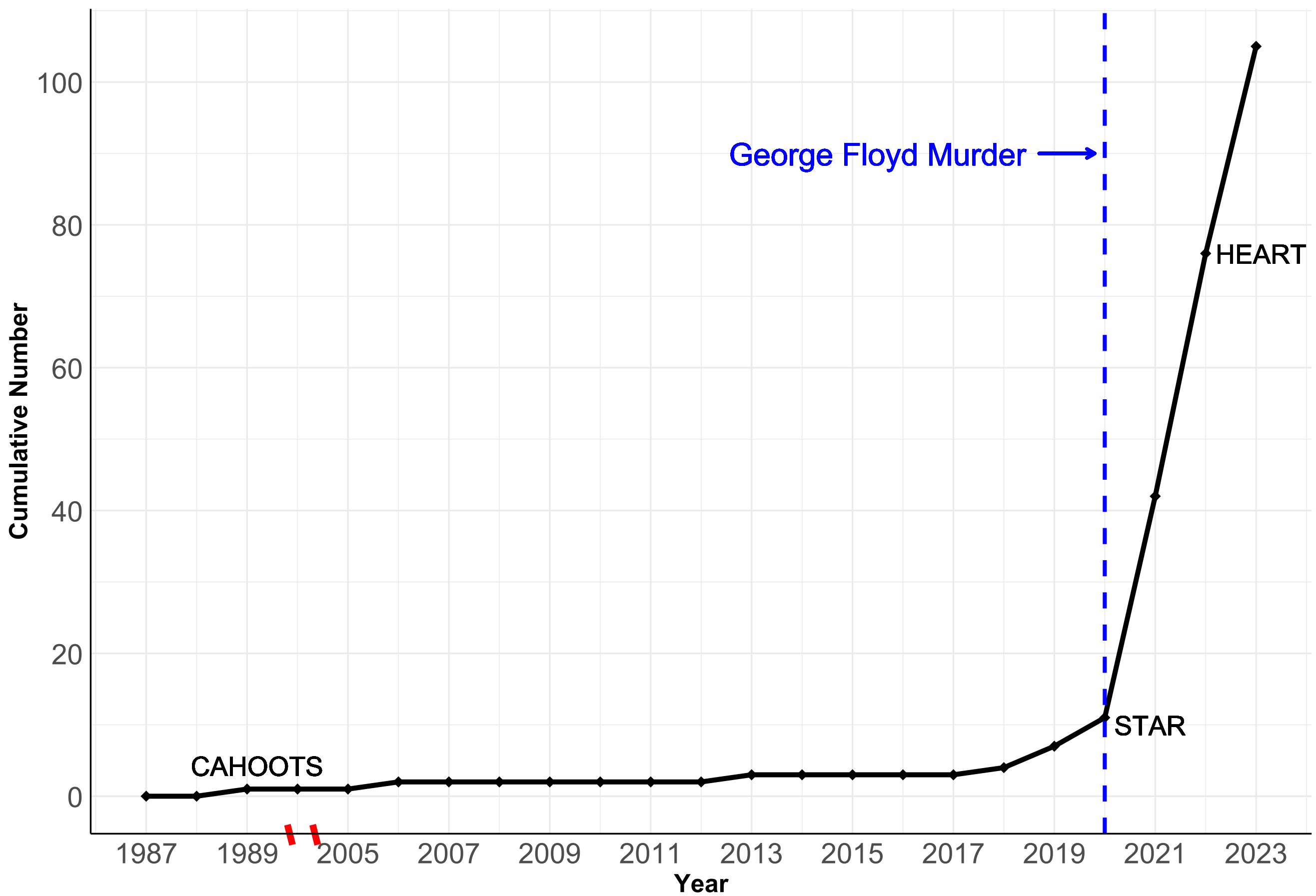
Trumbull, W. N. (1990). Who has standing in cost-benefit analysis? *Journal of Policy Analysis and Management 9*(2), 201–218.

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.

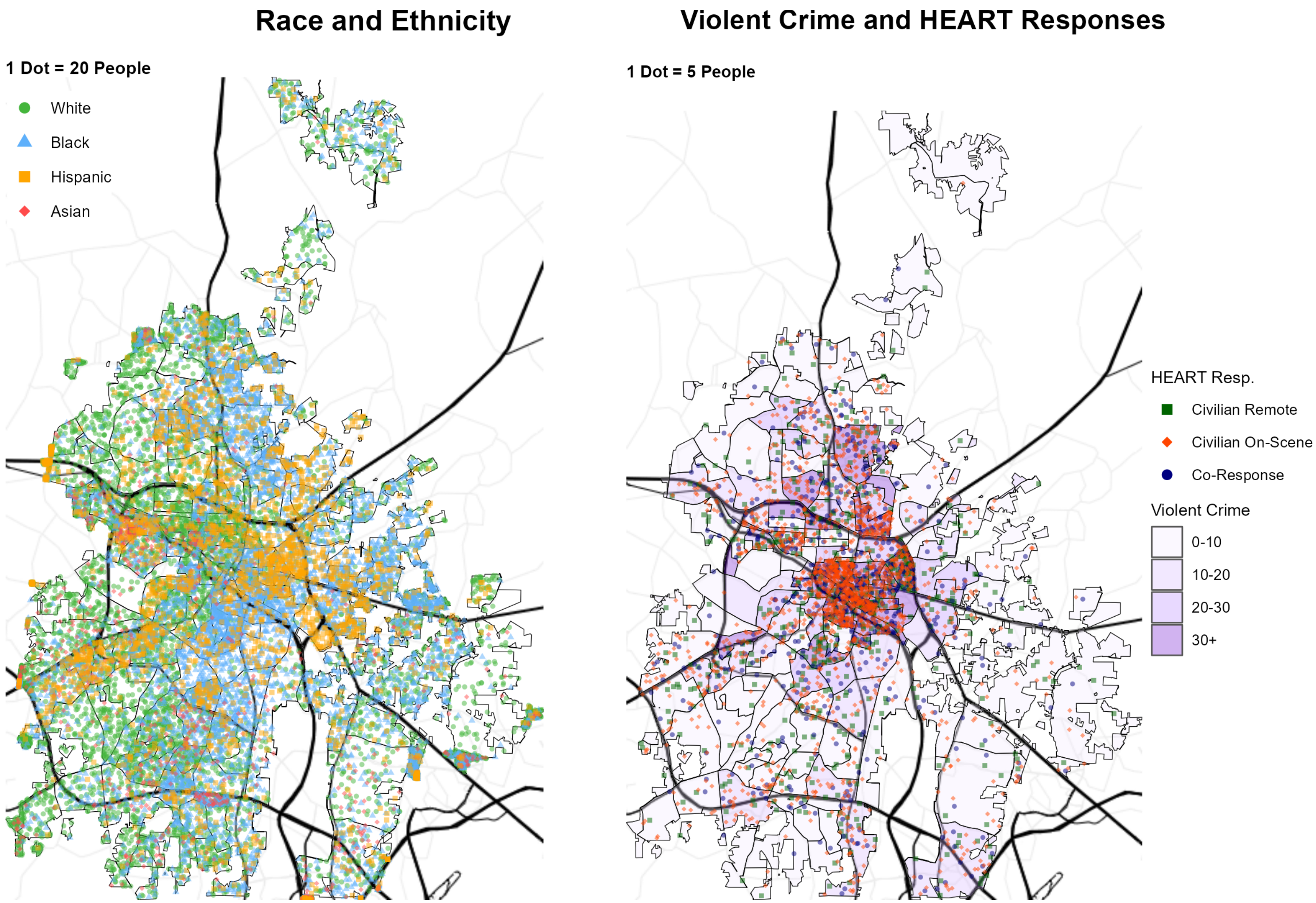
Yegen, E. (2021). Do institutional investors mitigate social costs of privatization? evidence from prisons. Technical report.

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.

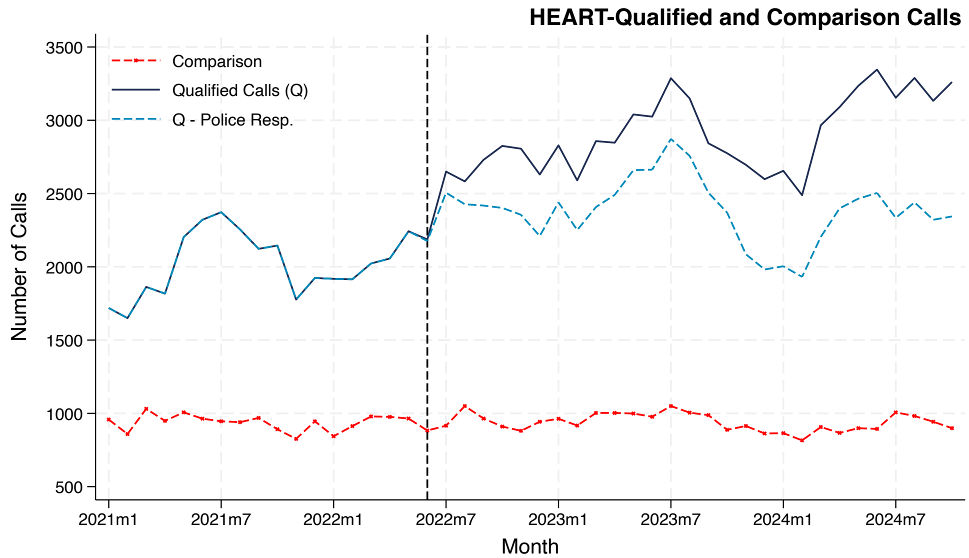
Figure 2: Demographics, Crime, and HEART Responses in Durham



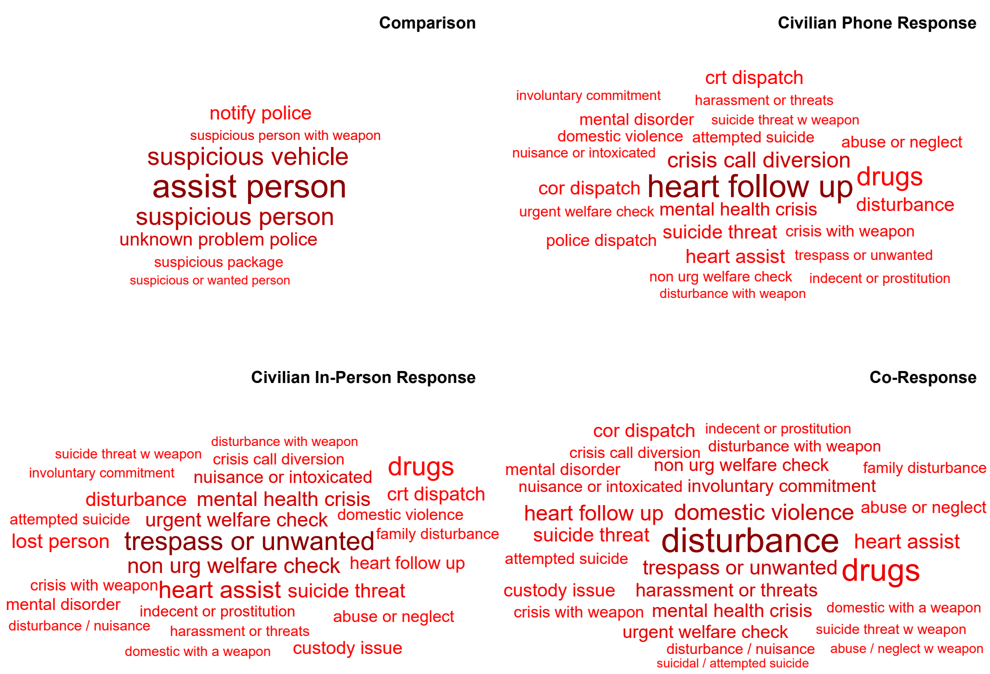
Notes: The left panel displays the racial and ethnic composition of Durham's population using 2019 ACS data at the census block group level, with each dot representing 20 residents. The right panel presents violent crime rates per 1,000 residents in 2019 and HEART response locations.

Figure 3: HEART-Qualified Cases, HEART Responses, and Police Responses

(a) Trends Over Time

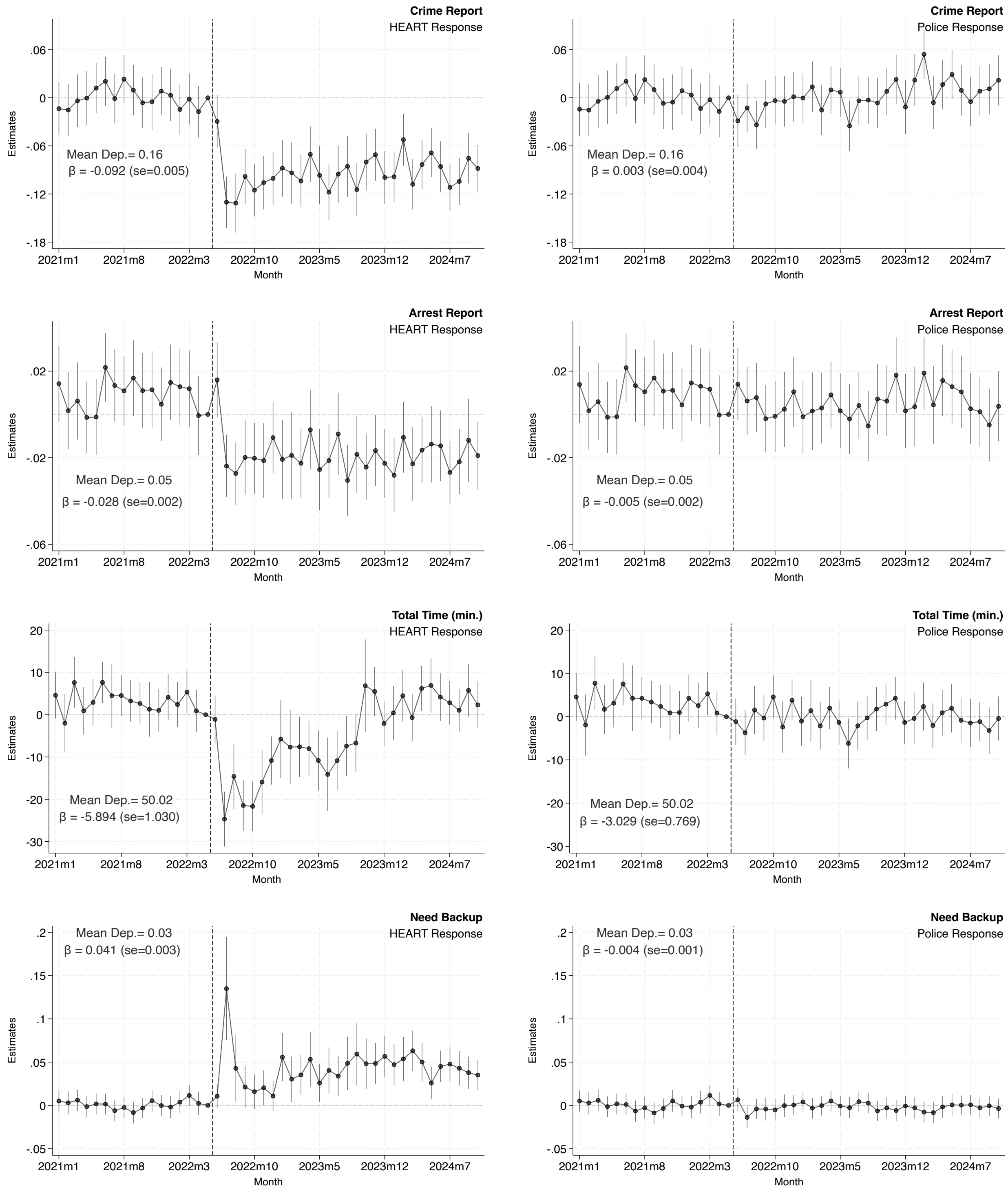


(b) Word Clouds Associated with Nature of the Call



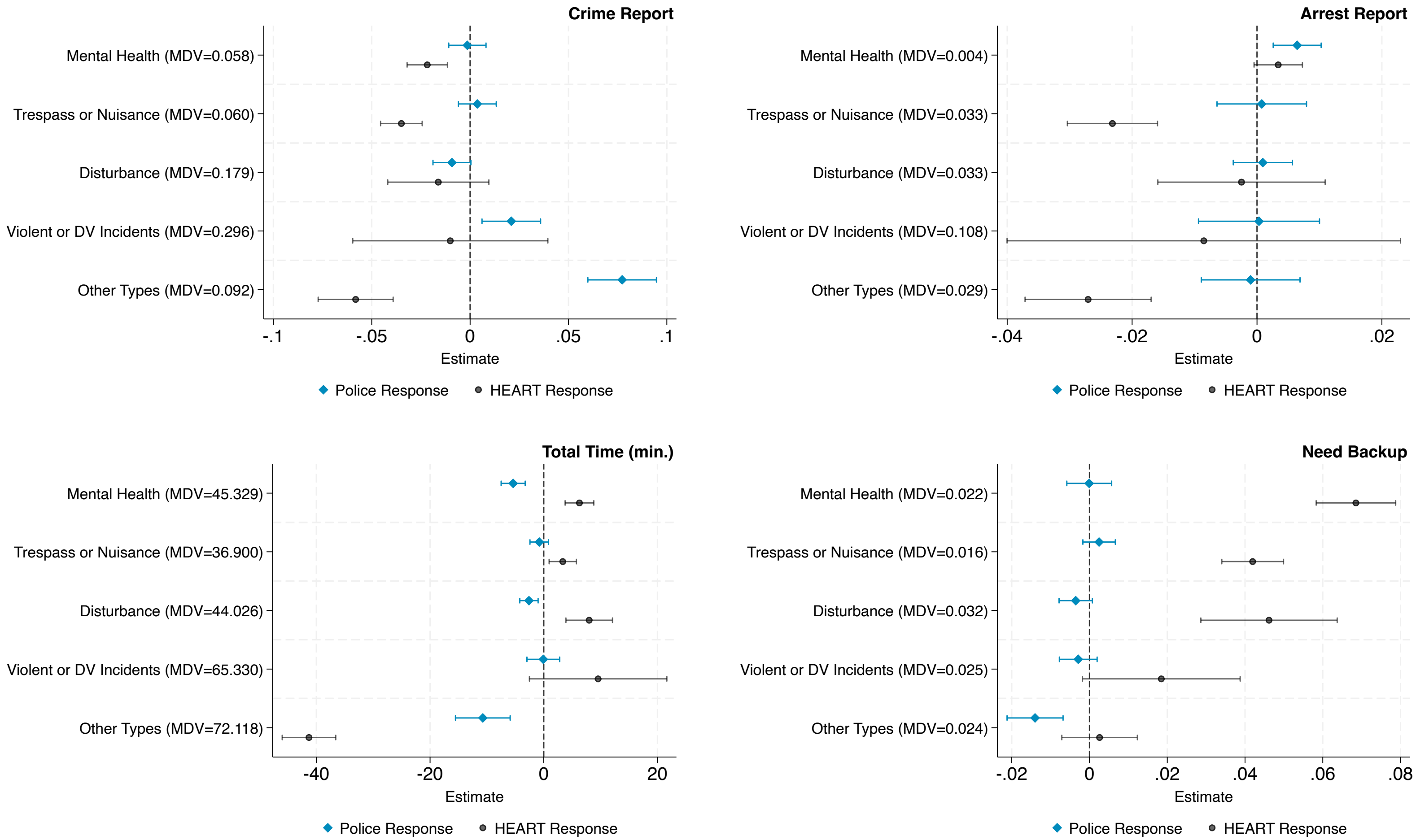
Notes: The top figure shows monthly call trends from 2021 to 2024, distinguishing HEART-qualified cases and comparison calls. The qualified cases fall under civilian phone response, civilian in-person response, and police–civilian co-response. The bottom panel presents word clouds of the call type descriptions for each response model, based on the nature codes associated with 911 dispatches. The font size reflects the relative frequency of each call type within a given category of emergency response.

Figure 4: Impact of HEART Program on Incident Outcomes



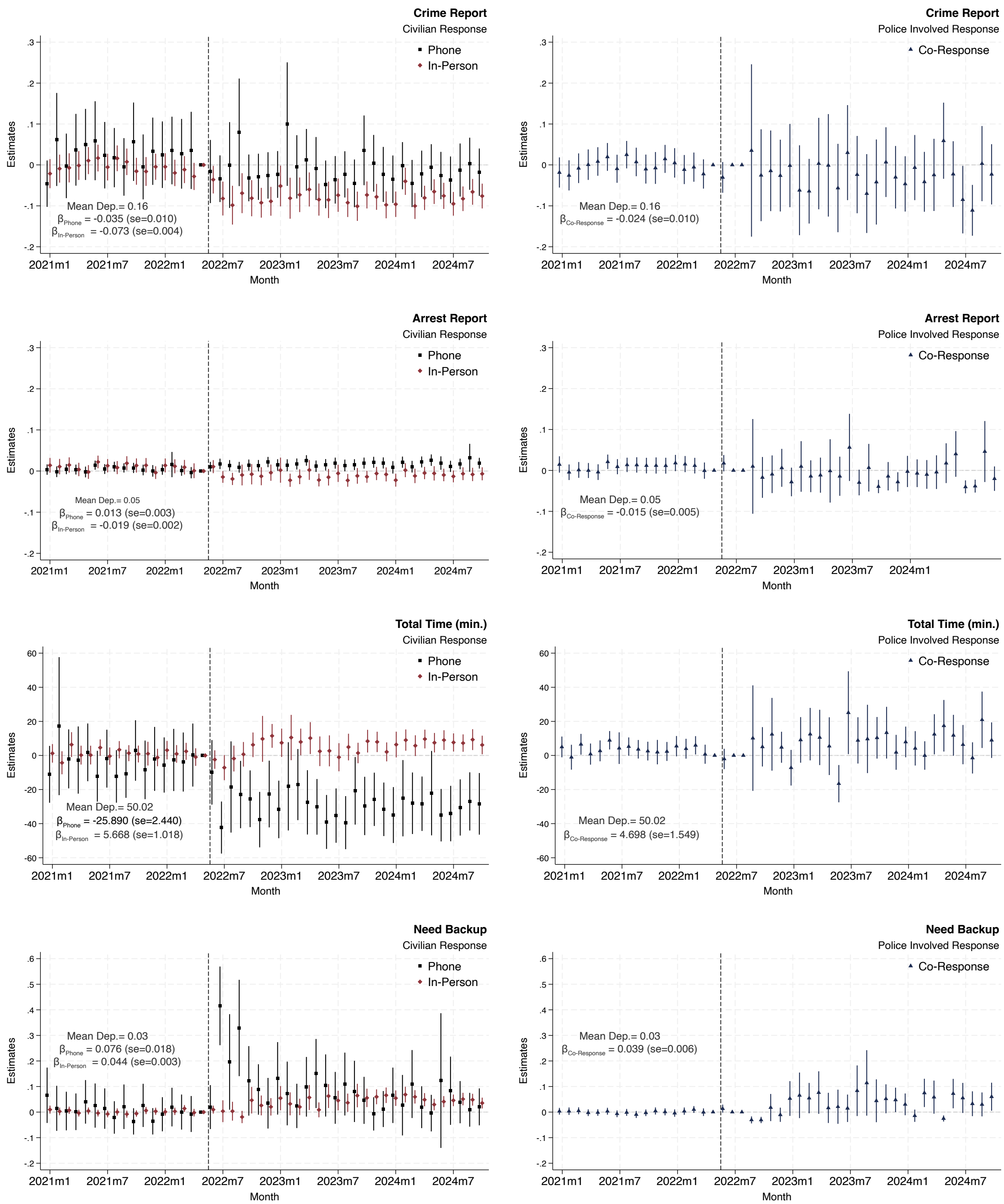
Notes: These figures present the impact of the HEART program on incident outcomes, including crime, arrests, response time, and requests for backup. They present DiD estimates for HEART (left-hand side) and police (right-hand side) responses with coefficients and 95% confidence intervals. The mean of the dependent variable for qualified 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: Heterogeneity by Nature of Call



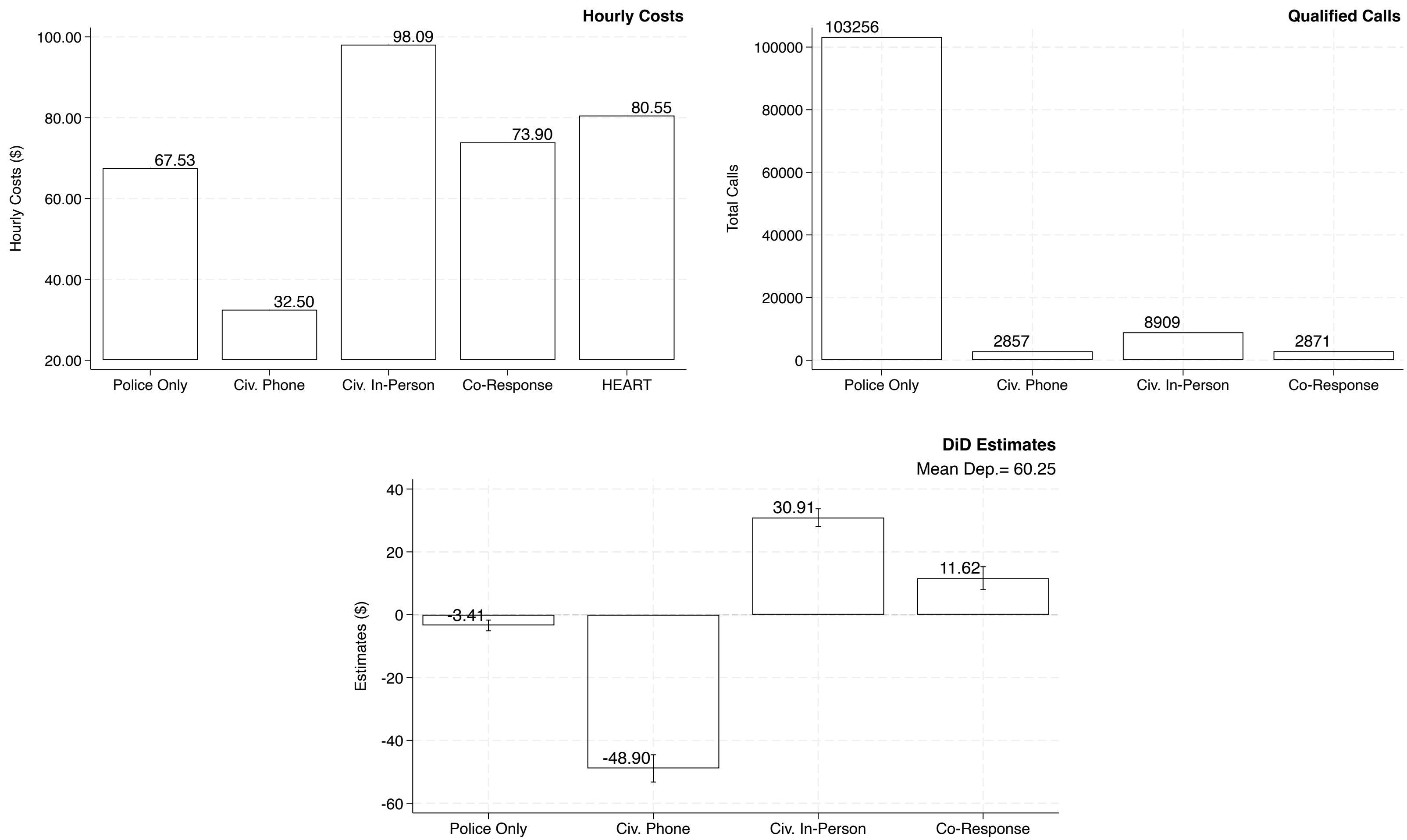
Notes: These figures present the impact of the HEART program on key outcomes, including crimes, arrests, dispatch time, and clearing time, disaggregated by the nature of the call: mental health, trespass or nuisance, disturbance, violent or domestic violence incidents, and other types of calls. The "other 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 6: Impact on Incident Outcomes by HEART Program



Notes: These figures present the impact of the Crisis Call Diversion (civilian phone response), Community Response Teams (civilian in-person response), and Co-Response programs on crimes, arrests, dipatch times, and clearing times. Civilian phone response cases involve mental health crises, suicide threats, or crisis calls diverted from police dispatch. The civilian in-person 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 qualified 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.

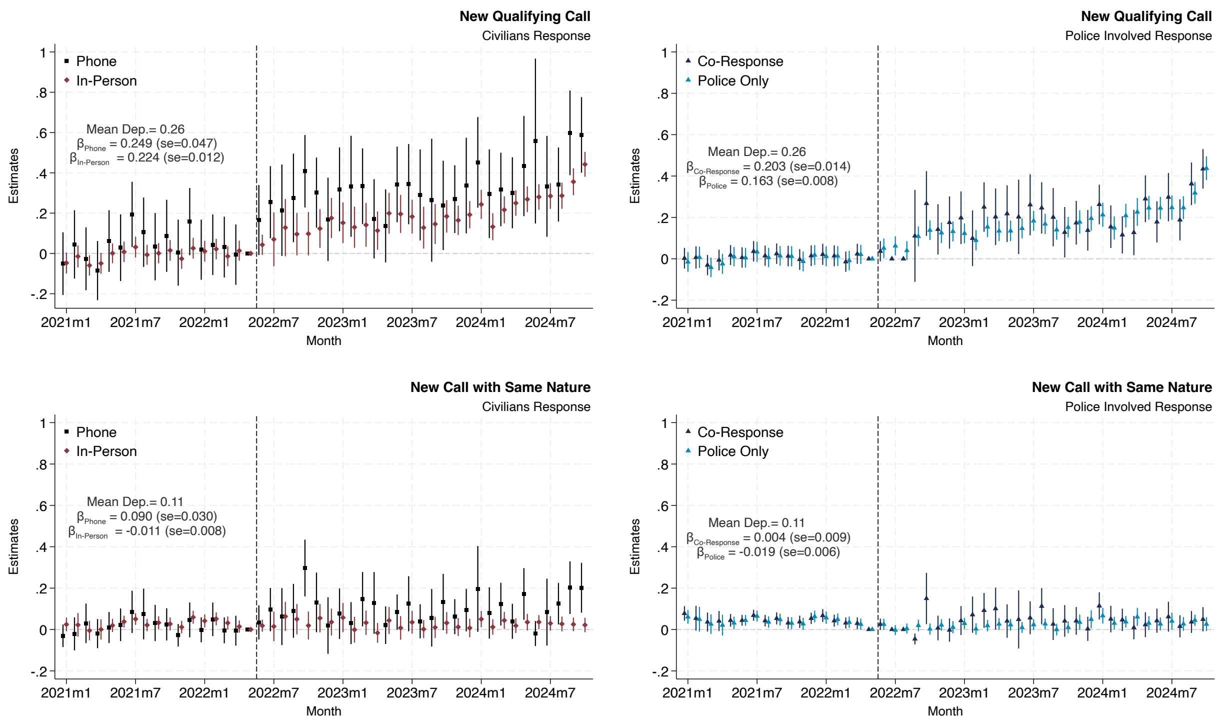
Figure 7: Impact of HEART on Labor Costs



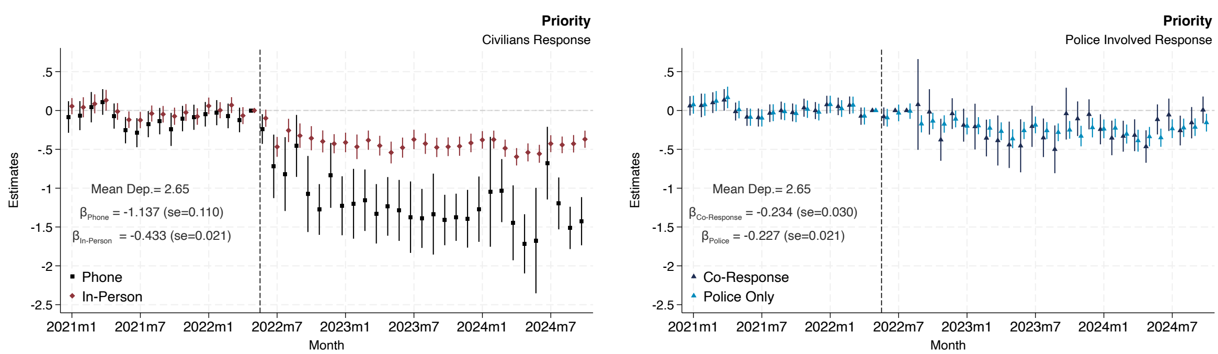
Notes: The top-left figure shows the average hourly wage for each different response model (civilian phone response, civilian in-person response, and co-response), calculated as annual salary divided by 52 weeks and 40 hours per week, based on position descriptions from the program and Durham's public safety agencies. The estimates reflect mean wages within each job category. Police wages reflect two officers per response; civilian in-person response wages cover clinicians, EMTs, and safety specialists; and co-response wages include a police officer paired with a mental health professional. The top-right figure presents the total number of HEART-qualified calls by response type. The bottom figure presents the program'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 qualified 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 8: Does HEART Improve Trust in 911?

A) Impact of Each HEART Program on Subsequent Calls of the Same Nature and Eligibility

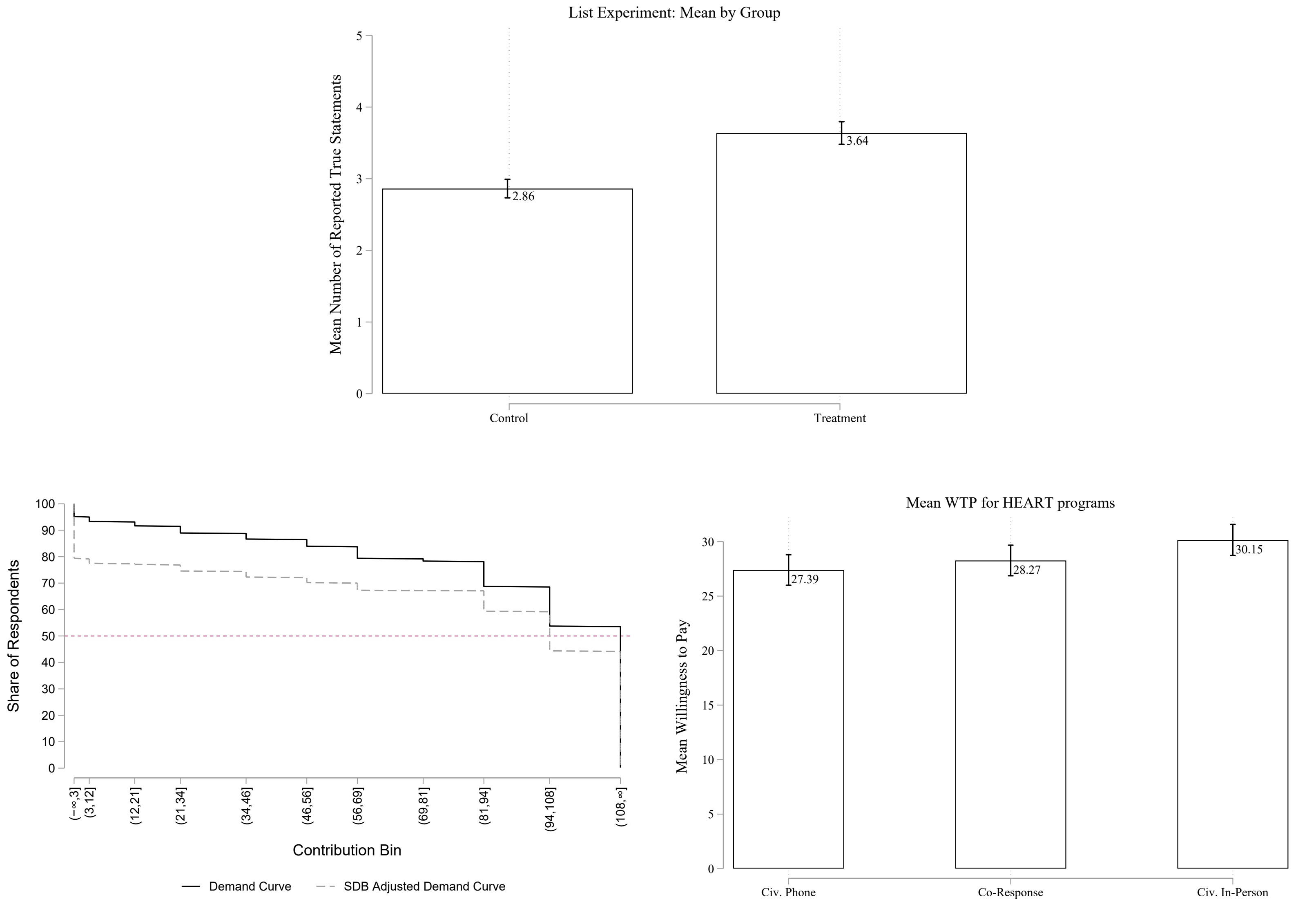


B) The Relationship between HEART introduction and Call Priority



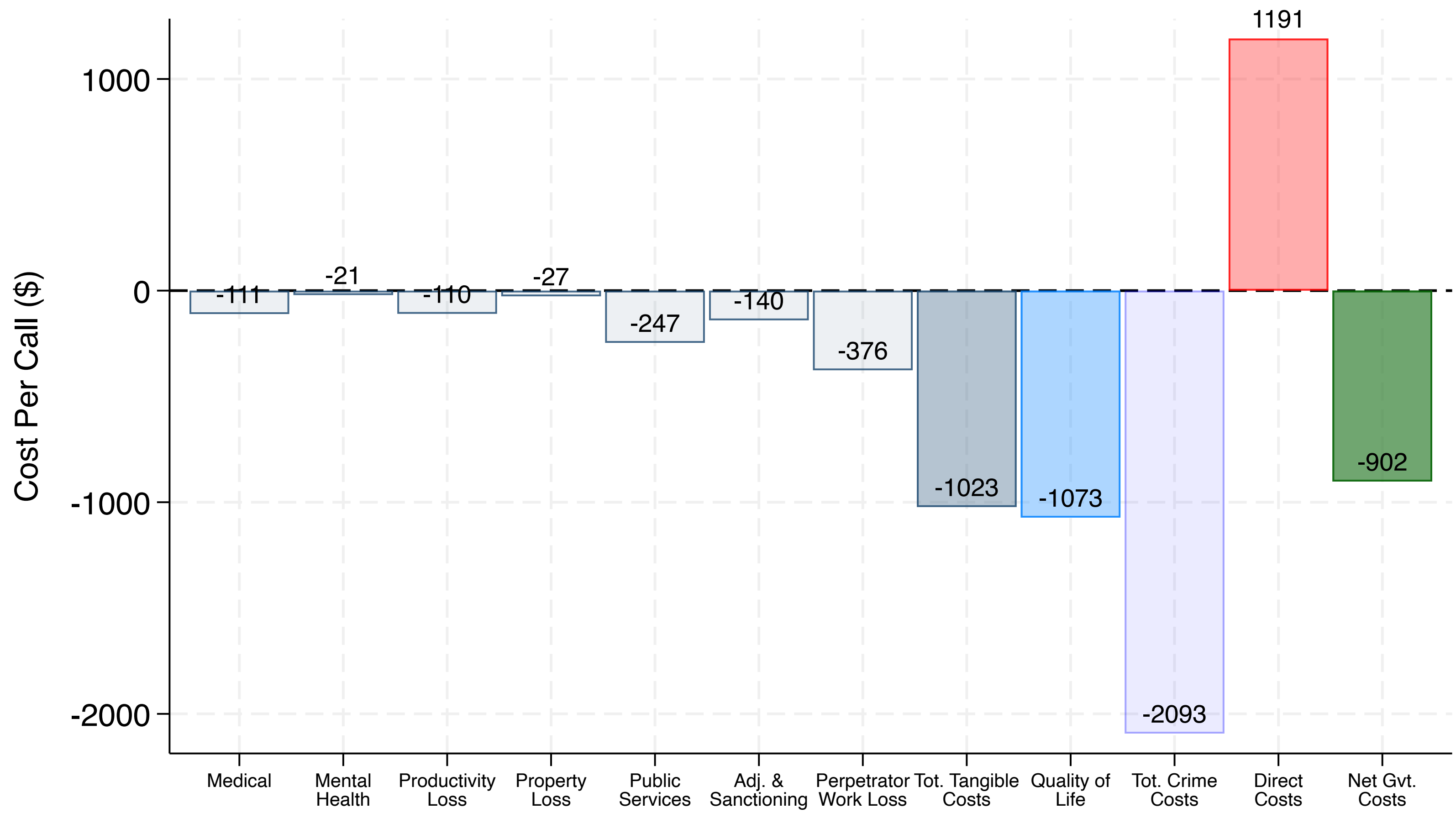
Notes: Panel A presents the impact of HEART on subsequent calls of the same nature and eligibility. We report DiD estimates with 95% confidence intervals, evaluating the effects of police, HEART, Crisis Call Diversion (civilian phone response), Community Response Teams (civilian in-person response), and Co-Response on the priority score. The mean of the dependent variable for qualified cases in the pre-period is included for reference. Regressions control for time and location factors, including day-of-week, hour, call priority, month–year, and beat fixed effects, with standaTherd errors clustered at the address level. Panel B presents the relationship between HEART introduction and 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 (civilian phone response), Community Response Teams (civilian in-person response), and Co-Response on the priority score. The mean of the dependent variable for qualified 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 9: Willingness to Pay for HEART



Notes: These figures present estimates of willingness to pay (WTP) for the HEART program. The top panel displays results from the list experiment, comparing responses between the treatment and control groups; the difference between the two bars indicates the share of respondents who truthfully revealed their WTP. The bottom-left panel shows the overall demand curve and the demand curve corrected for social desirability bias. The bottom-right panel reports average WTP by program type: civilian phone response, civilian in-person response, and co-response. The bottom panels report means with 95% confidence intervals.

Figure 10: 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 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 the program’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) Comparison Police Resp.	(3) Qualify Police Resp.	(4) Qualify HEART Resp.	(5) Qualify (4)-(3) Coeff	(6) Qualify (4)-(3) p-value
A) Call Characteristics						
Share of Asian	0.039	0.041	0.039	0.035	0.000	0.533
Share of Black	0.413	0.408	0.416	0.412	-0.004	0.092
Share of Hispanic	0.177	0.173	0.178	0.183	-0.002	0.103
Share of 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 Back-Up	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
Use of Force (X 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-qualified, police response, and HEART response cases. Column (1) reports the mean for the full sample. Columns (2)-–(4) show the means by case type: comparison cases, HEART-qualified cases receiving police responses, and qualified cases receiving HEART responses. Column (5) presents the difference in means between the program and police responses among qualified 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 Program on Crimes and Arrests by Type of Offense

	(1) Violent Crime	(2) Arrest for Violent Crime	(3) Property Crime	(4) Arrest for Property Crime	(5) Other Crime	(6) Arrest for Other Crime
HEART	-0.0389***	-0.0151***	-0.00839***	-0.00348***	-0.0585***	-0.0116***
	(0.00225)	(0.00110)	(0.00123)	(0.000599)	(0.00398)	(0.00170)
Police	-0.00818***	-0.0128***	-0.00404***	-0.00314***	0.00646**	-0.00460***
	(0.00206)	(0.00112)	(0.00120)	(0.000596)	(0.00308)	(0.00152)
Mean Dep.	0.08	0.03	0.01	0.01	0.09	0.02
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Treatment Dummy	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
p-value:HEART=Police	0.00	0.00	0.00	0.17	0.00	0.00
Observations	161054	161054	161054	161054	161054	161054

Notes: This table reports the impact of the HEART program on crime reports and arrests, disaggregated by offense type. All regressions include fixed effects for day of week, hour, call priority, month–year, and police beat. Standard errors are clustered at the address level. Reported p-values test the null hypothesis that the HEART and police response coefficients are equal for HEART-qualified cases. The mean of the dependent variable in the pre-period is included for reference. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: Does HEART Lead to Escalation?

	(1) Use of Force (X 100)	(2) Any New Call	(3) New Call 1-31 Days	(4) New Call > 31 Days	(5) Any New Call for Violent Crime	(6) New Call for Violent Crime 1-31 Days	(7) New Call for Violent Crime > 31 Days
HEART	-0.0472	0.0183***	-0.00966	0.0425***	-0.00616***	-0.00412***	-0.00134
	(0.0431)	(0.00550)	(0.00929)	(0.00774)	(0.00214)	(0.00138)	(0.00101)
Police	-0.0377	0.0282***	-0.0117	0.0319***	-0.00401**	-0.00318***	-0.000417
	(0.0308)	(0.00416)	(0.00732)	(0.00600)	(0.00177)	(0.00115)	(0.000876)
Mean Dep.	0.13	0.96	0.43	0.21	0.03	0.01	0.01
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Treatment Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value:HEART=Police	0.77	0.01	0.82	0.06	0.12	0.31	0.15
Observations	161054	161054	161054	161054	161054	161054	161054

Notes: This table reports the impact of the HEART program on the likelihood of escalation. Our measures of escalation include: (1) an indicator for whether the incident involved use of force (coded as 1 if yes, 0 otherwise, and multiplied by 100 for interpretability), (2) whether the beneficiary placed any subsequent 911 calls, and (3) whether any future call led to a crime report involving a violent offense. All regressions include fixed effects for day of week, hour, call priority, month–year, and police beat. Standard errors are clustered at the address level. Reported p-values test the null hypothesis that the HEART and police response coefficients are equal for HEART-qualified cases. The mean of the dependent variable in the pre-period is included for reference. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Impact of HEART on Arrestee Profiles

	(1) Black	(2) White	(3) Hispanic	(4) Male	(5) Female	(6) $<$ 24 yo	(7) 24-39 yo	(8) $>$ 39 yo
HEART	-0.0242***	-0.00296***	-0.00131**	-0.0222***	-0.00741***	-0.00520***	-0.0147***	-0.00954***
	(0.00200)	(0.000913)	(0.000664)	(0.00202)	(0.00106)	(0.000884)	(0.00154)	(0.00127)
Police	-0.00447**	-0.000307	-0.0000183	-0.00356**	-0.00127	-0.00105	-0.00269*	-0.000934
	(0.00175)	(0.000823)	(0.000653)	(0.00175)	(0.000928)	(0.000888)	(0.00140)	(0.00113)
Mean Dep.	0.04	0.01	0.01	0.04	0.01	0.01	0.03	0.01
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Treatment Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value:HEART=Police	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	161054	161054	161054	161054	161054	161054	161054	161054

Notes: This table reports the impact of the HEART program on arrestee characteristics: race, ethnicity, gender, and age. All regressions include fixed effects for day of week, hour, call priority, month–year, and police beat. Standard errors are clustered at the address level. Reported p-values test the null hypothesis that the HEART and police response coefficients are equal for HEART-qualified cases. The mean of the dependent variable in the pre-period is included for reference. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Contents

Supplementary Materials

A Additional Specifications

A.1 Impact of HEART from TSLS

Estimation Strategy As a complementary analysis, we adapt the approach of Ouss and Stevenson (2023) and estimate the effect of a HEART response on various outcomes with TSLS. We leverage the program's launch and its targeted qualification criteria to construct an instrument based on the variation in the policy's effect on qualified calls. We recover the TSLS estimates by estimating the following equations:

$$HEART_{it} = \alpha_F + Post_t \cdot Qualify_i \cdot \beta^{FS} + Qualify_i \cdot \lambda_F + \gamma_t^F + X'_{ic} \cdot \delta_F + \epsilon_{ict}^F \tag{9}$$

$$Y_{ict} = \alpha_H + HEART_{it} \cdot \beta^H + Qualify_i \cdot \lambda_H + \gamma_t^H + X'_{ic} \cdot \delta_H + \epsilon_{ict}^H \tag{10}$$

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

Following Imbens and Angrist (1994), the key identifying assumptions for β^H are relevance, independence, monotonicity and the exclusion restriction. Relevance requires that the instrument significantly predict the likelihood of a HEART response, meaning that qualified calls post-introduction must be likelier to have received the intervention. Independence assumes that the potential outcomes are independent of the HEART assignment, conditional on the instrument, indicating that the trends in outcomes would have been similar for qualified and comparison 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 that a call would receive a HEART response.

Relevance and Balance Tests Figure A.9 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 qualified calls rises sharply and remains elevated. The coefficient on the interaction between the post-introduction and qualification indicators is 0.147 ($SE = 0.003$), with a nF-statistic of 1,874.9, confirming a strong first stage. For the period prior to the program's implementation, the trends in outcomes for qualified and comparison calls are flat and parallel, supporting the identification strategy. The sharp post-introduction divergence indicates the program's substantial impact on the program's response rates, reinforcing the strength of the instrument.

Table A.3 assesses the validity of the independence assumption by testing whether the instrument, $Post \times Qualified$, 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 the program's responses are slightly less likely in areas with a larger Asian population share but exhibit no strong patterns otherwise. Column (2) tests the instrument, $Post \times Qualify$, 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 the program's responses (Column (1)) is statistically insignificant ($p = 0.125$), indicating no correlation with call characteristics. The result from the joint F-test for the instrument (Column (2)) is also insignificant ($p = 0.162$), supporting the assumption that the instrument is balanced.

Monotonicity We expect monotonicity to hold mechanically as HEART did not exist in the pre-period. However, as a test of the monotonicity assumption, we examine whether the first-stage effect of qualifying for the HEART program remains nonnegative across a range of subsamples defined by observable call characteristics and neighborhood context. Table A.4 reports estimates from regressions of treatment receipt on the $Post \times Qualify$ indicator. Column (1) shows the baseline estimate for the full sample, with a coefficient of 0.147 (SE = 0.0034), indicating a strong and statistically significant first-stage relationship. Columns (2) through (5) disaggregate the sample by the share of South Asian, Black, White, and Hispanic residents in the caller's census block group (above median), while Columns (6) and (7) split the sample by local unemployment rates. Across all specifications, the first-stage estimate remains positive and highly significant, ranging from 0.137 to 0.157. These results provide strong support for the monotonicity assumption, as qualifying for the program consistently predicts an increased likelihood of receiving treatment across diverse demographic and economic contexts.

A.2 Impact of Each Program According to TSLS

Estimation Strategy We estimate the impact of each HEART response type—H_1 (civilian phone), H_2 (civilian in-person), and H_3 (co-response)—on key call-level outcomes using a two-stage least squares (TSLS) framework. We exploit the qualification criteria and the introduction of HEART in June 2022 as sources of variation. Specifically, we instrument each realized HEART response using the interaction between the post-period indicator and response-type-specific qualification dummies.

We estimate the following first-stage equation:

$$H_{1,it} = Z_{it}^1 \cdot \beta_1^{FS} + \lambda_1 \cdot Qualify_i^1 + X_{ic}' \cdot \delta_1 + \gamma_t^1 + \delta_c^1 + \epsilon_{it}^1 \tag{11}$$

$$H_{2,it} = Z_{it}^2 \cdot \beta_2^{FS} + \lambda_2 \cdot Qualify_i^2 + X_{ic}' \cdot \delta_2 + \gamma_t^2 + \delta_c^2 + \epsilon_{it}^2 \tag{12}$$

$$H_{3,it} = Z_{it}^3 \cdot \beta_3^{FS} + \lambda_3 \cdot Qualify_i^3 + X_{ic}' \cdot \delta_3 + \gamma_t^3 + \delta_c^3 + \epsilon_{it}^3 \tag{13}$$

and the corresponding second-stage specification:

$$Y_{ict} = \alpha + \widehat{H}_{1it} \cdot \beta_1 + \widehat{H}_{2it} \cdot \beta_2 + \widehat{H}_{3it} \cdot \beta_3 + X_{ic}' \cdot \Gamma + \gamma_t + \delta_c + \epsilon_{ict} \tag{14}$$

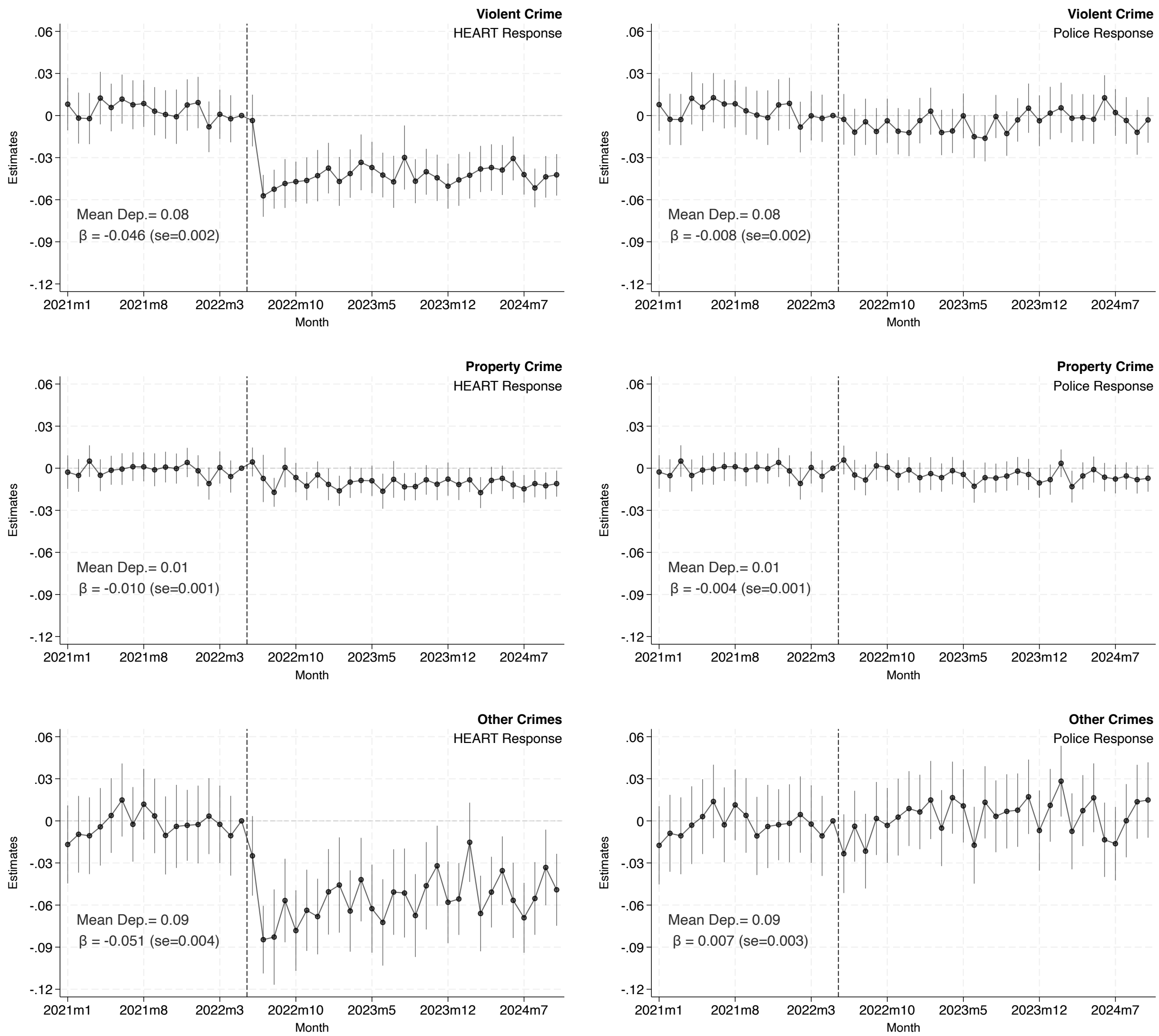
where $j \in \{1, 2, 3\}$ indexes the three response types. We use $Post_t$, an indicator for calls occurring after the launch of HEART, and its interaction with the qualification indicators $Qualify_i^j$ as instruments to capture variation in exposure to each response type, such that $Z_{it}^j = Post_t \cdot Qualify_i^j$. The estimation includes fixed effects for day-of-week, hour-of-day, call priority, geographic beat, and calendar month of dispatch, as well as a vector of call-level covariates X_{ic}. Standard errors are clustered at the address level.

First-Stage Table A.6 presents first-stage estimates showing how each instrument predicts assignment to each HEART response type. The diagonal elements of the table, where each instrument is used to predict its corresponding response type (e.g., instrument for civilian phone predicting civilian phone), are large in magnitude and highly statistically significant. These diagonal coefficients range from 0.017 to 0.172, indicating that qualified cases are substantially more likely to receive the intended response following the implementation of HEART.

In contrast, the off-diagonal elements capture how each instrument predicts the other response types. These coefficients are consistently negative and significant, albeit smaller in magnitude, suggesting that eligibility for one HEART model reduces the likelihood of receiving the others. For instance, the instrument for civilian in-person decreases the probability of receiving a civilian phone response by 4.26 pp, and vice versa. This pattern reflects that each case is qualified for only one HEART response type, and eligibility for one model systematically crowds out assignment to the others.

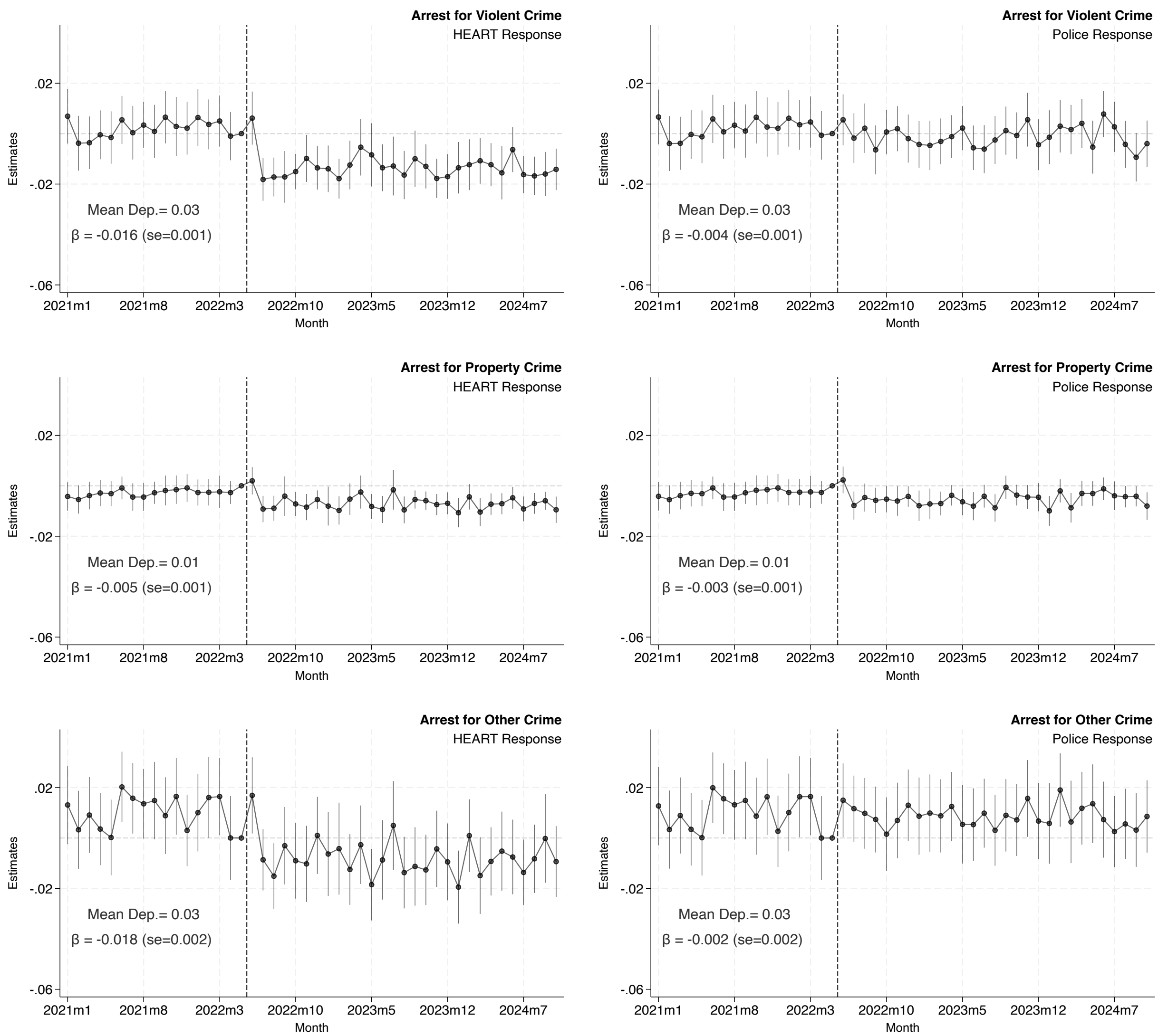
B Additional Figures and Tables

Figure A.1: Impact of HEART Program on Crime by Type of Offense



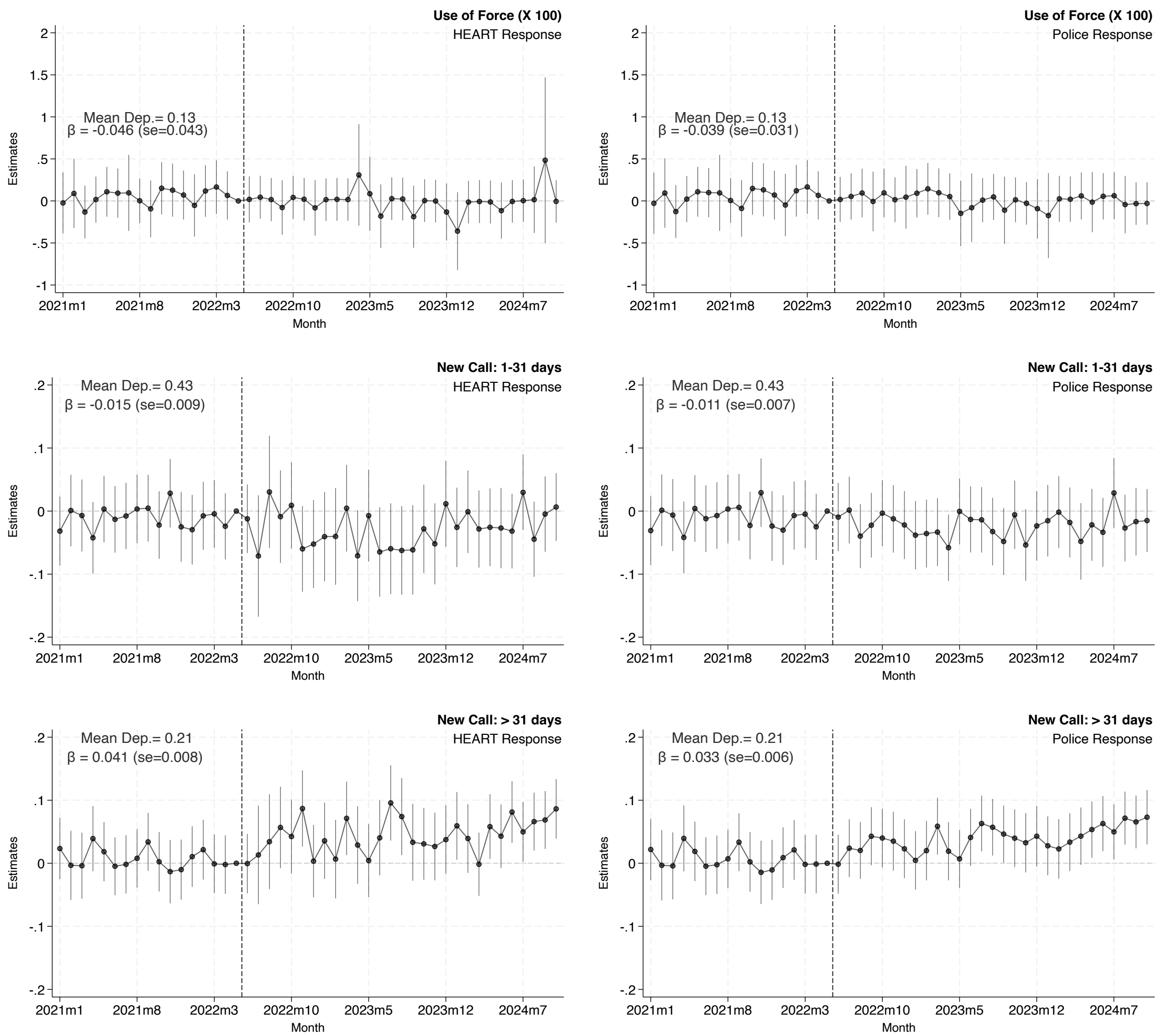
Notes: These figures present the impact of the HEART program on crimes. They present DiD estimates for HEART (left-hand side) and police (right-hand side) responses with coefficients and 95% confidence intervals. The mean of the dependent variable for qualified cases in the pre-period is included for 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: Impact of HEART Program on Arrests by Type of Offense



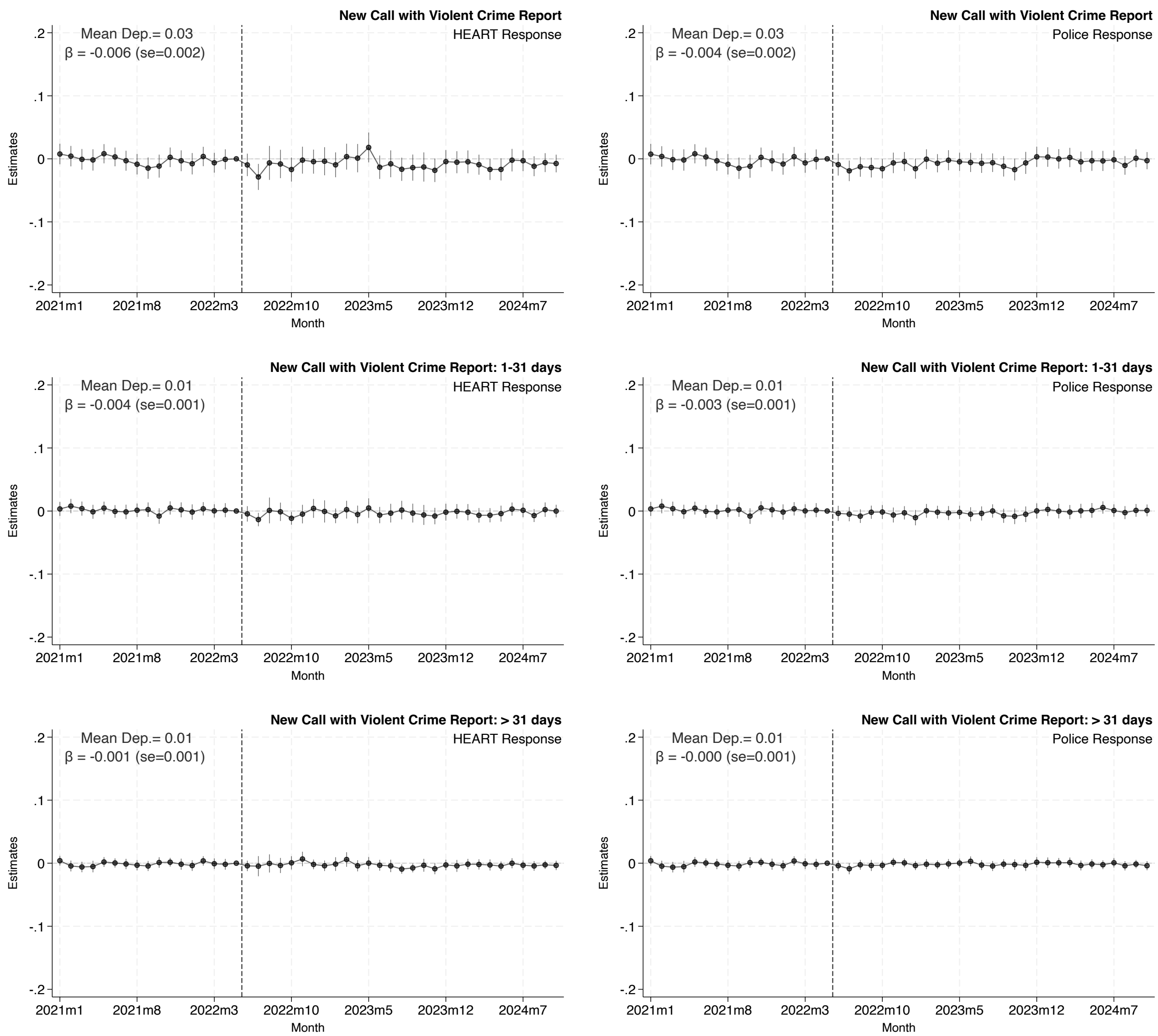
Notes: These figures present the impact of the HEART program on arrests. They present DiD estimates for HEART (left-hand side) and police (right-hand side) responses with coefficients and 95% confidence intervals. The mean of the dependent variable for qualified cases in the pre-period is included for 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.3: Does HEART Lead to Escalation?



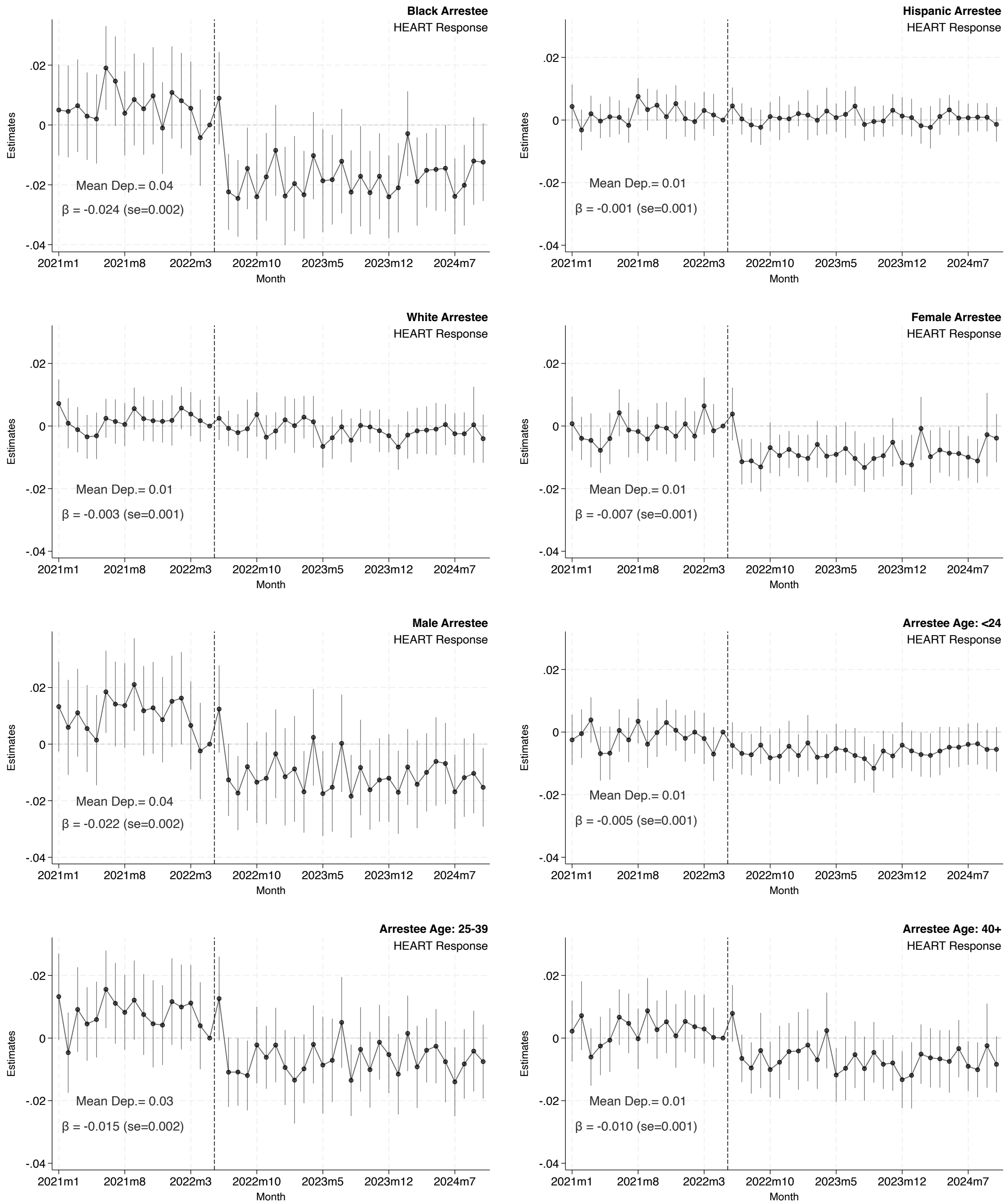
Notes: These figures present the impact of the HEART program on the likelihood of escaltion. Our measures of escalation include: (1) an indicator for whether the incident involved use of force (coded as 1 if yes, 0 otherwise, and multiplied by 100 for interpretability), (2) whether the beneficiary placed any subsequent 911 calls, and (3) whether any future call led to a crime report involving a violent offense. All regressions include fixed effects for day of week, hour, The figures present DiD estimates for HEART (left-hand side) and police (right-hand side) responses with coefficients and 95% confidence intervals. The mean of the dependent variable for qualified 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.4: Does HEART Lead to Escalation? (Continued)



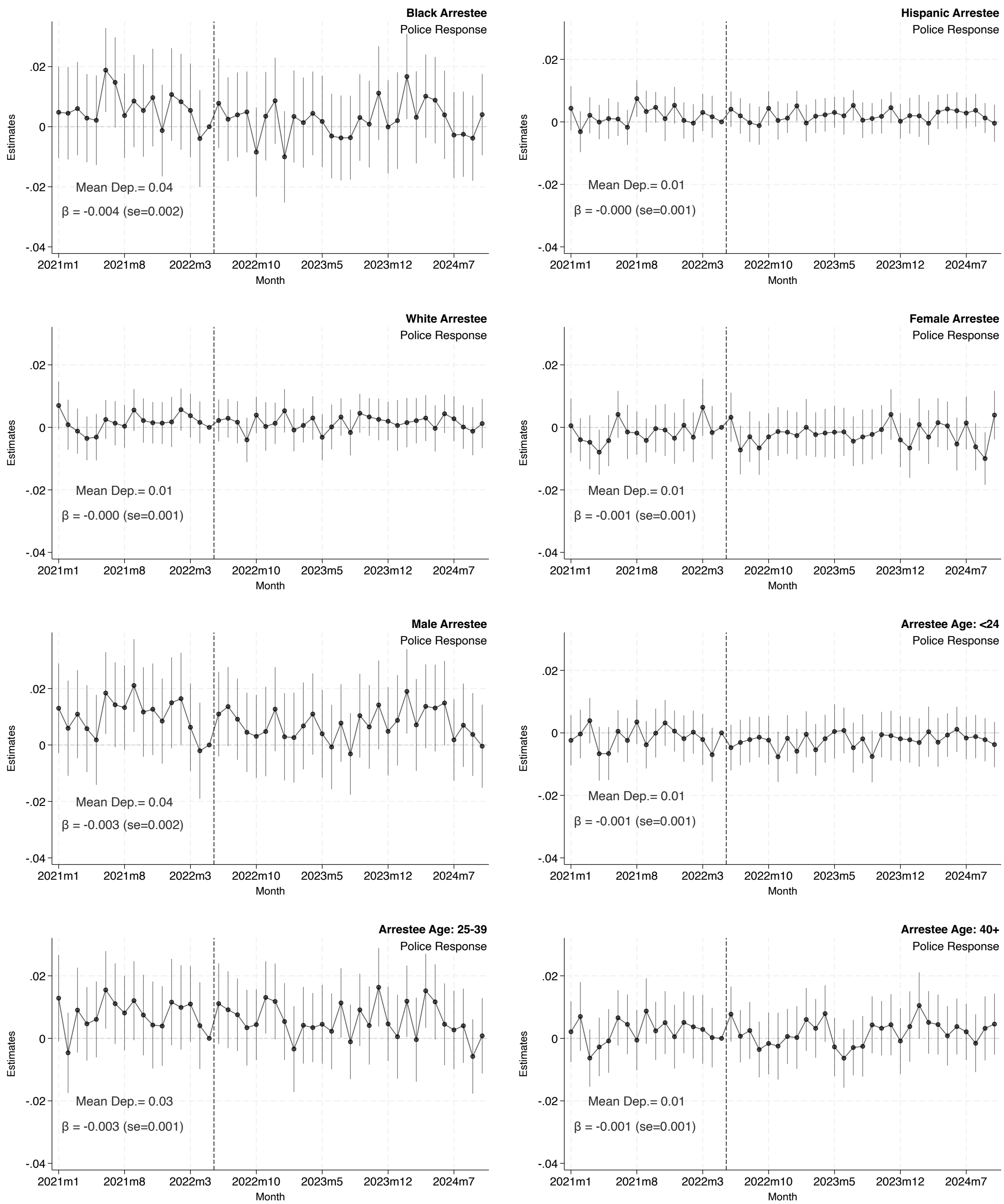
Notes: These figures present the impact of the HEART program on the likelihood of escaltion. Our measures of escalation include: (1) an indicator for whether the incident involved use of force (coded as 1 if yes, 0 otherwise, and multiplied by 100 for interpretability), (2) whether the beneficiary placed any subsequent 911 calls, and (3) whether any future call led to a crime report involving a violent offense. All regressions include fixed effects for day of week, hour,

Figure A.5: Impact of HEART Response on Arrestee Characteristics



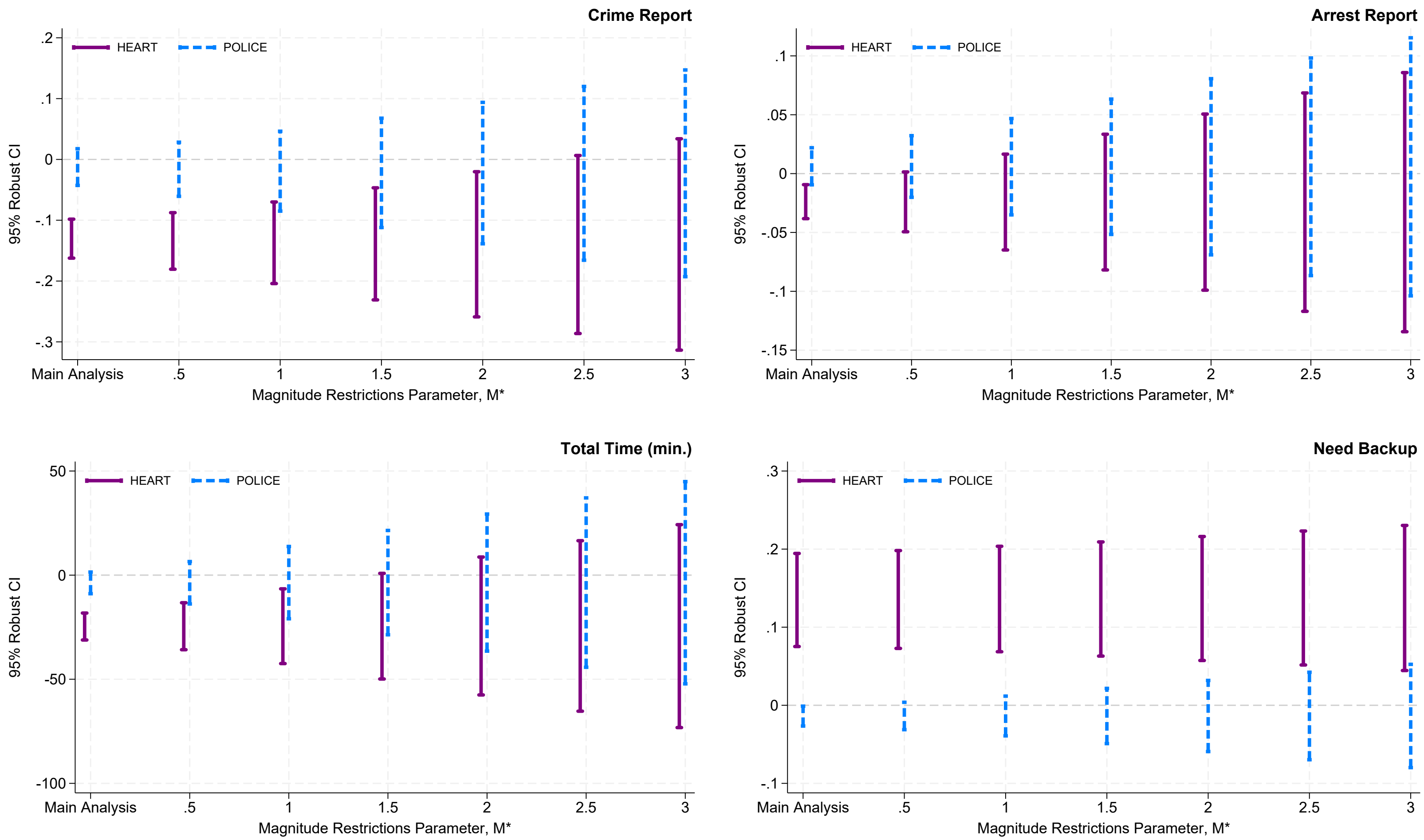
Notes: These figures present the impact of HEART responses on arrestee characteristics. It presents DiD estimates for police and HEART responses, with coefficients and 95% confidence intervals. The mean of the dependent variable for qualified cases in the pre-period is included for reference. Regressions 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 Police Response on Arrestee Characteristics



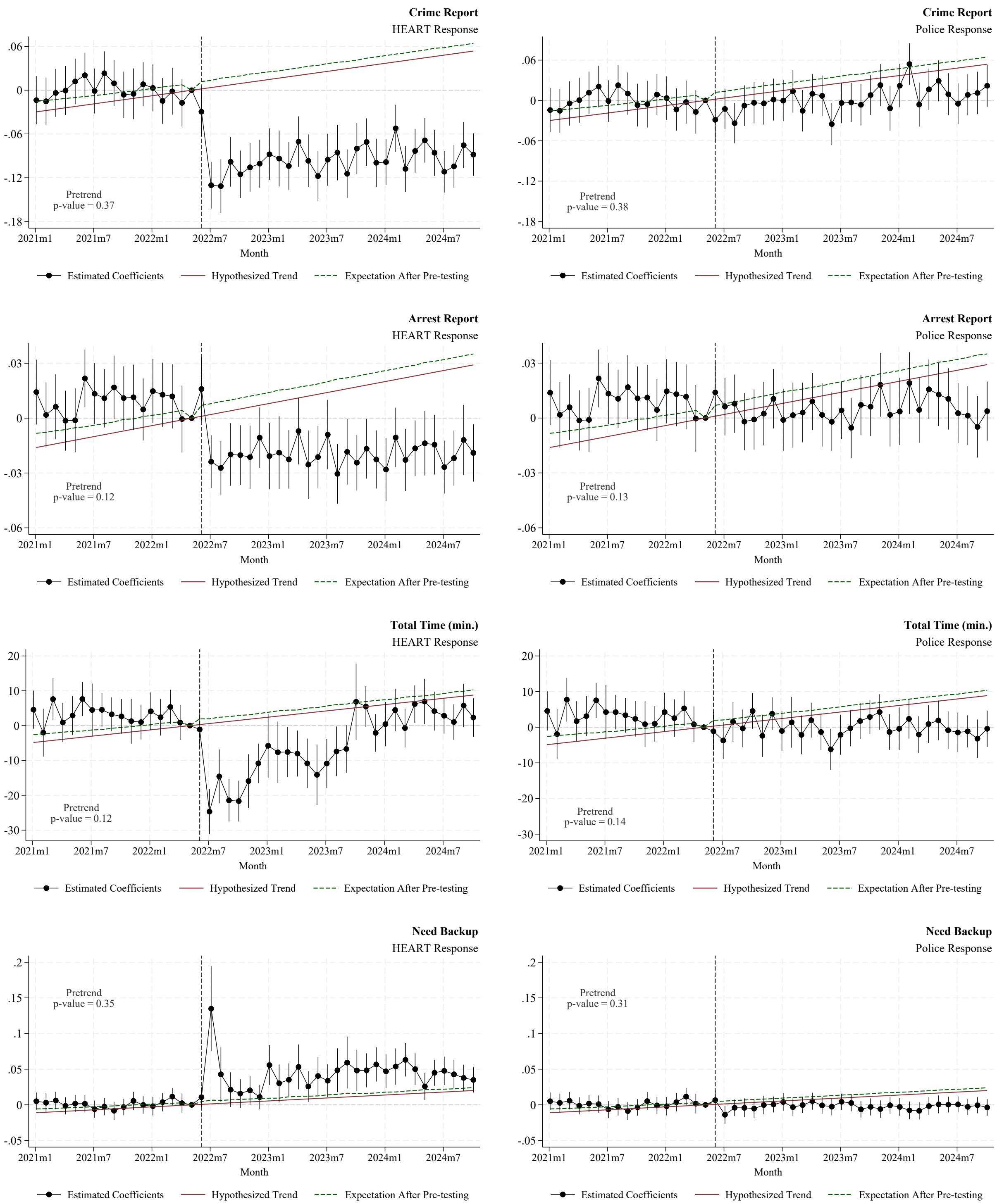
Notes: These figures present the impact of police responses on arrestee characteristics for HEART-qualified cases. It presents DiD estimates for police and HEART responses, with coefficients and 95% confidence intervals. The mean of the dependent variable for qualified cases in the pre-period is included 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 Relative Magnitudes for DiD



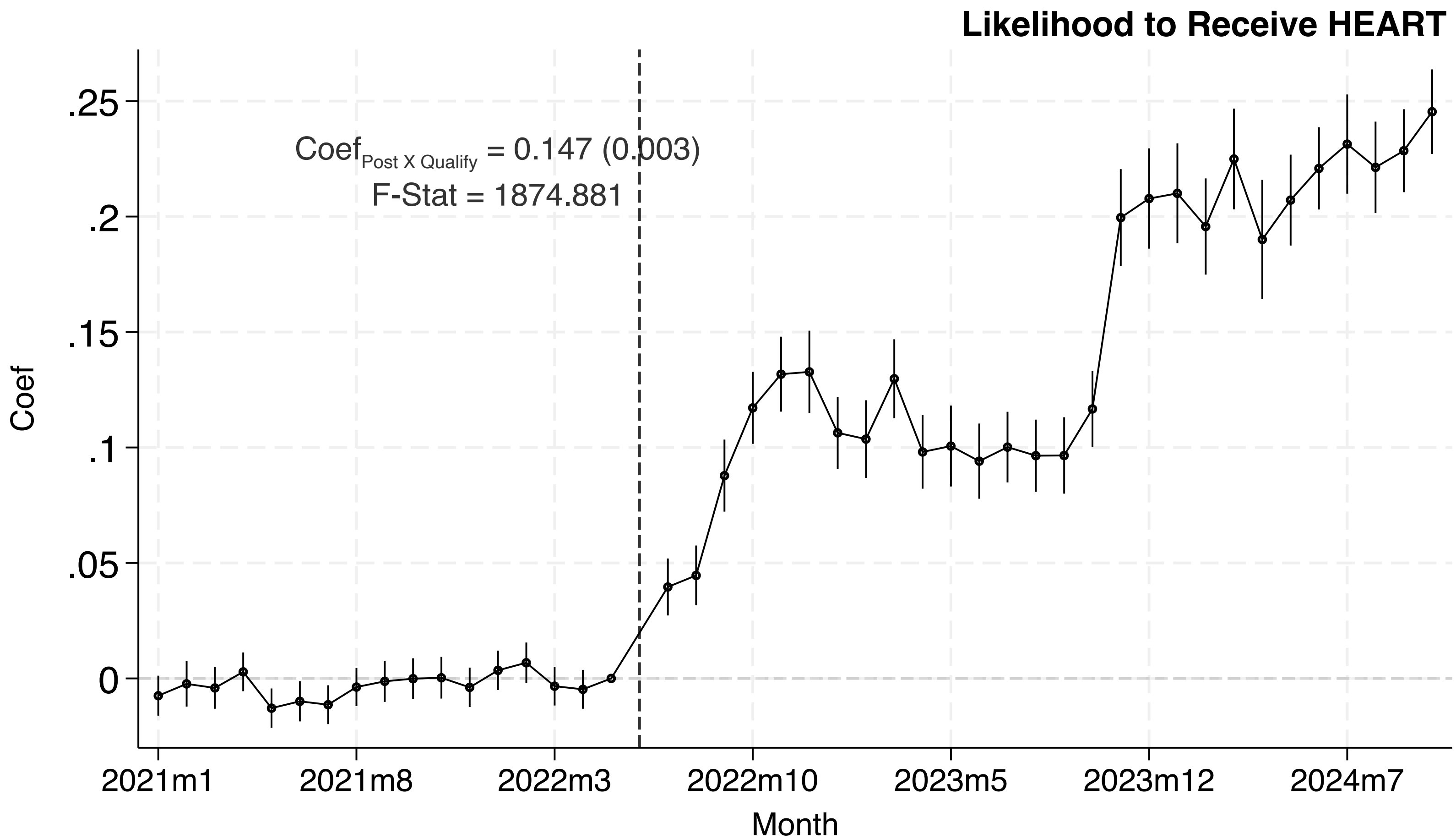
Notes: These figures present results from various specification checks on the impact of HEART on various outcomes. The figure presents the bounds on relative magnitudes associated with a 95% confidence interval from Rambachan and Roth (2023).

Figure A.8: Event Study and Hypothesized Trends



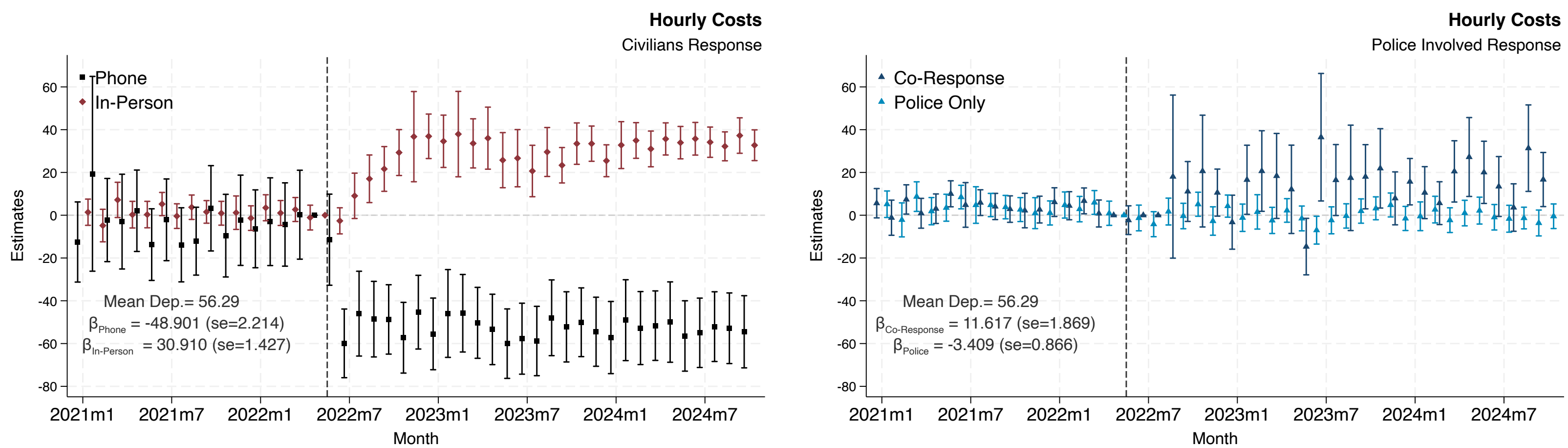
Notes: These figures present results from various specification checks on the impact of HEART on various outcomes. The figures plot potential violations of parallel trends based on Roth (2022). We report the event-study coefficients from Figure 4 and 95% confidence intervals, with standard errors clustered 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 the event-study coefficients if deviations existed but were undetectable under conventional methods.

Figure A.9: First Stage: Effect of HEART Implementation on HEART Response Rates



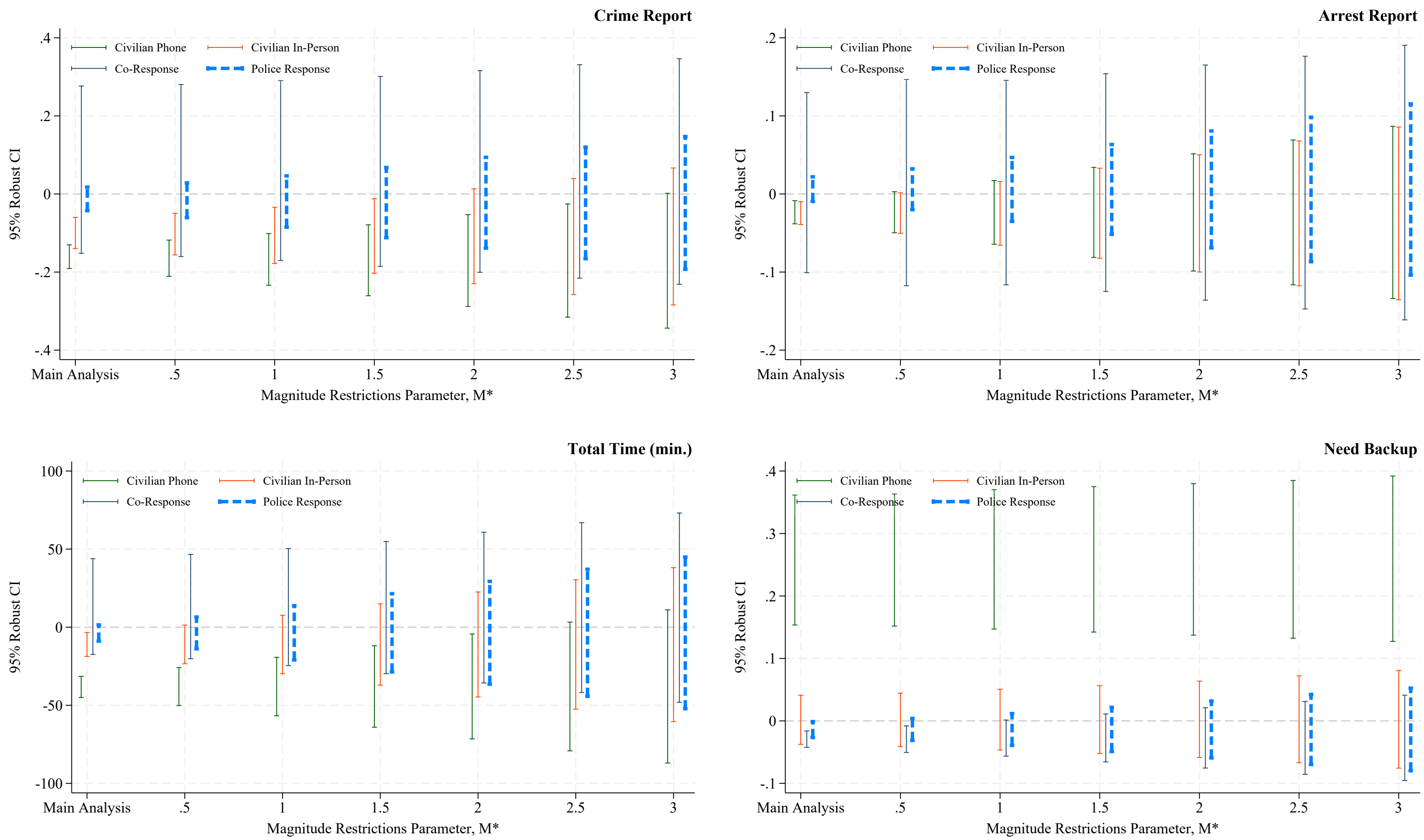
Notes: This figure shows results from the first-stage regression estimating the probability of a HEART response with the implementation of HEART as an instrument. The interaction between the time since introduction and an indicator that a case qualifies for a HEART response serves as the instrument, with unqualified calls as the control group. The regression controls for the covariates in Table A.3 and includes 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.

Figure A.10: Impact of HEART on Labor Costs



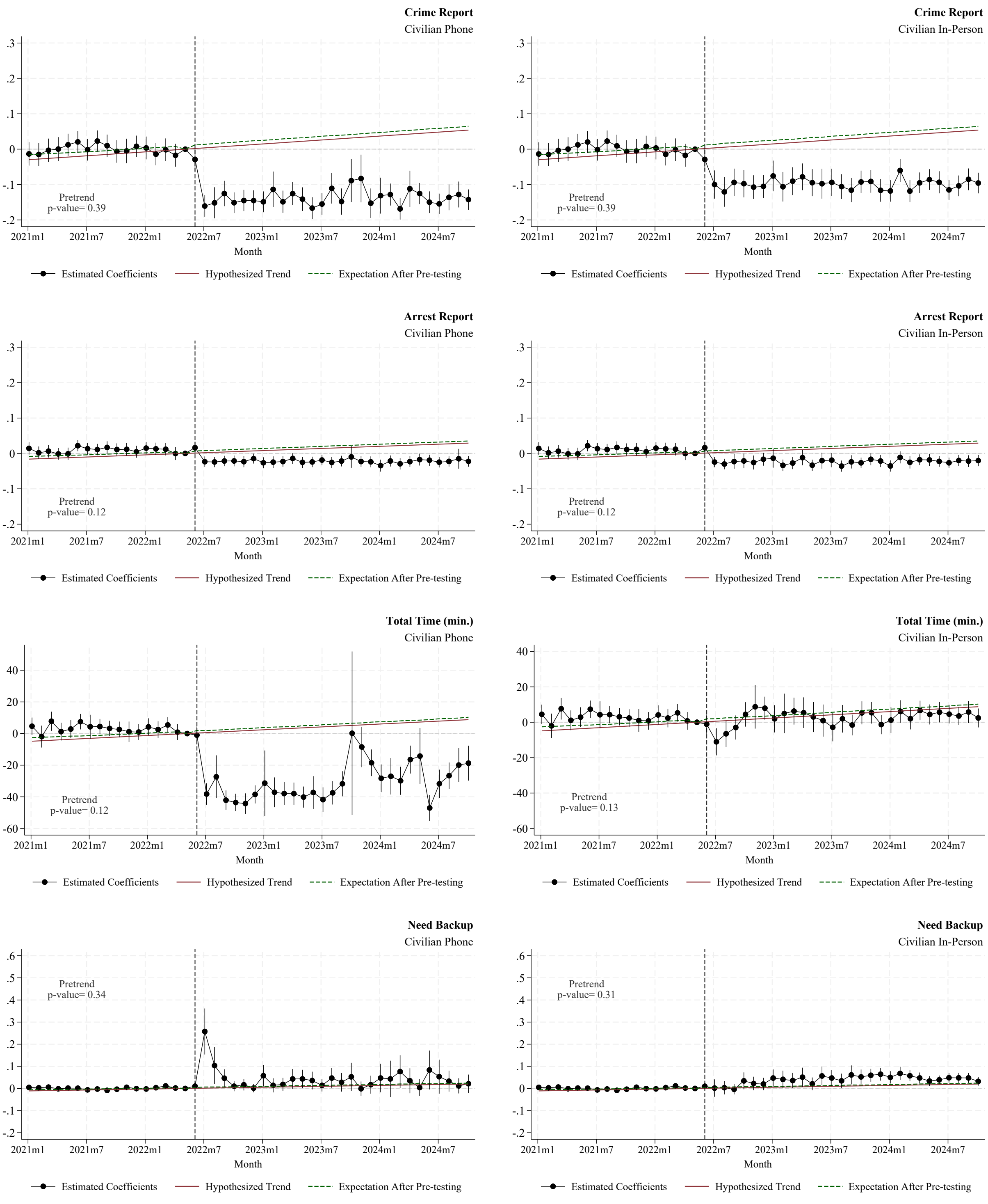
Notes: The figures present the impact of HEART on labor costs, which we calculate by multiplying the average hourly wage of responders for each response model by the time required to handle the calls. We report DiD estimates with 95% confidence intervals, evaluating the effects of civilian phone response, civilian in-person response, and co-response, or a police-only response on labor costs. The mean of the dependent variable for qualified 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.11: Bounds on Relative Magnitudes for DiD for Each HEART Program



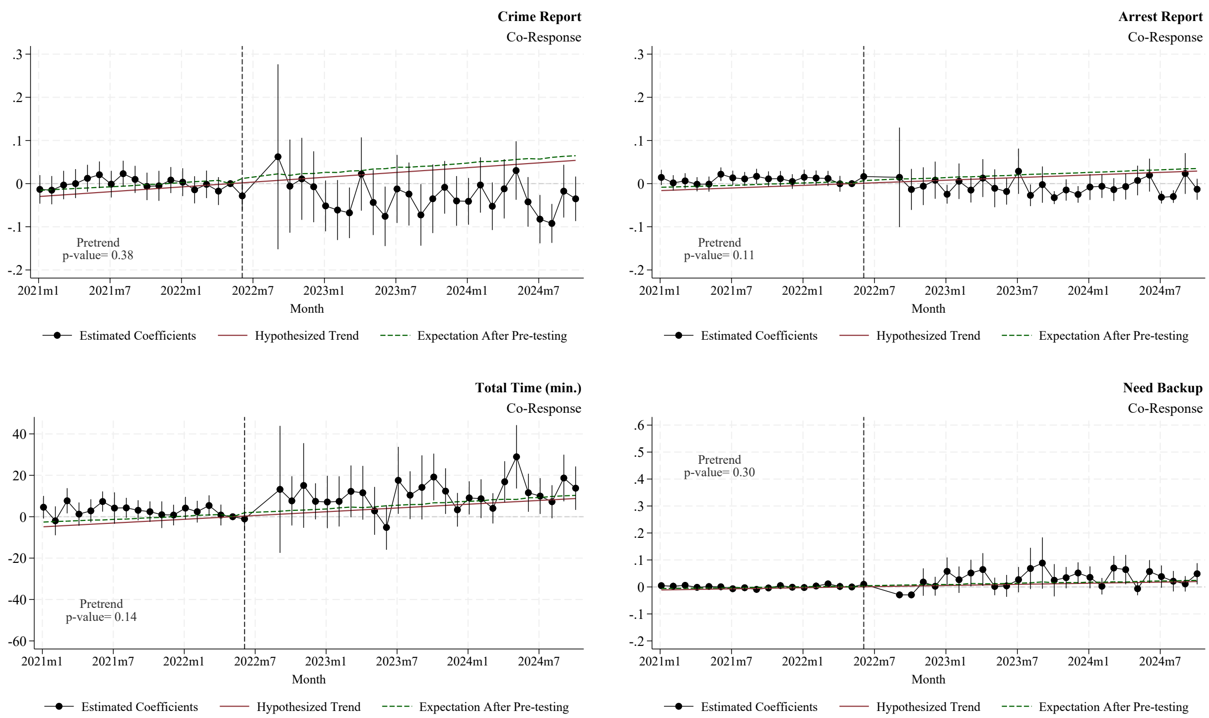
Notes: These figures present results from various specification checks on the impact of each HEART team on various outcomes. The figure presents the bounds on relative magnitudes associated with a 95% confidence interval from Rambachan and Roth (2023).

Figure A.12: Event Study and Hypothesized Trends for Civilian Responses



Notes: These figures present results from various specification checks on the impact of civilian 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 the event-study coefficients if deviations existed but were undetectable by conventional methods.

Figure A.13: Event Study and Hypothesized Trends for Co-Response



Notes: These figures present results from 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 the event-study coefficients if deviations existed but were undetectable by conventional methods.

Figure A.14: Iterated Multiple Price List Experiment

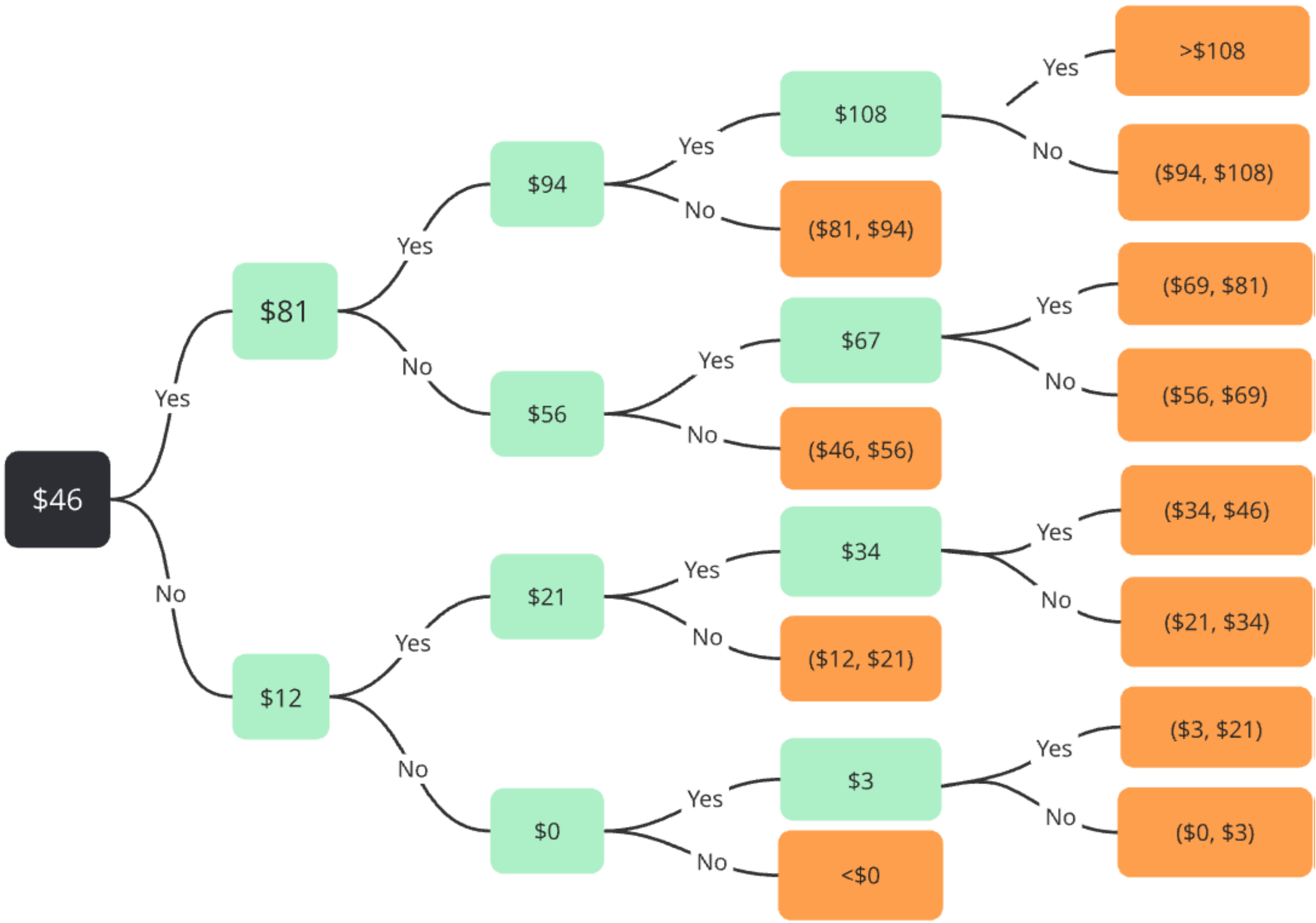


Table A.1: Contingent Valuation Starting Value Balance Tests

		Treatment Arm			
	All (1)	Starting Value \$21 (2)	Starting Value \$46 (3)	Starting Value \$81 (4)	p-value test (5)
Age	45.12	44.26	45.65	45.43	0.73
	(0.77)	(1.34)	(1.37)	(1.27)	
Female (%)	58.75	55.35	57.32	63.69	0.29
	(2.25)	(3.95)	(3.87)	(3.85)	
Race: White (%)	56.67	59.75	54.27	56.05	0.60
	(2.26)	(3.90)	(3.90)	(3.97)	
Race: Black (%)	31.46	28.93	33.54	31.85	0.67
	(2.12)	(3.61)	(3.70)	(3.73)	
Race: Other (%)	11.88	11.32	12.20	12.10	0.97
	(1.48)	(2.52)	(2.56)	(2.61)	
Hispanic (%)	9.79	9.43	9.76	10.19	0.97
	(1.36)	(2.33)	(2.32)	(2.42)	
College (%)	53.33	52.83	51.22	56.05	0.68
	(2.28)	(3.97)	(3.92)	(3.97)	
Married (%)	43.96	42.77	42.68	46.50	0.74
	(2.27)	(3.94)	(3.87)	(3.99)	
Employed (%)	69.17	70.44	66.46	70.70	0.65
	(2.11)	(3.63)	(3.70)	(3.64)	
Income	87727.27	90397.35	85691.82	87203.95	0.85
	(3483.28)	(5718.47)	(5932.72)	(6458.86)	
Homeowner (%)	49.38	50.31	48.17	49.68	0.92
	(2.28)	(3.98)	(3.91)	(4.00)	
Democrat (%)	52.92	47.17	59.15	52.23	0.10
	(2.28)	(3.97)	(3.85)	(4.00)	
Observations	480	159	164	157	

Notes: This table reports reduced-form results using the contingent valuation survey. Column (1) presents the overall mean. Column (2) presents the mean for the group that first conisdered voting for HEART at \$21. Columns (3) and (4) consider analogous means for groups that first considered voting for HEART for \$46 and \$81, respectively. Column (5) is a p-value from an F-test of joint equality across treatment conditions.

Table A.2: Contingent Valuation List Treatment Balance Tests

	Treatment Arm			
	All (1)	Control (2)	Treated (3)	p-value test (4)
Age	45.12 (0.77)	45.18 (1.07)	45.06 (1.09)	0.94
Female (%)	58.75 (2.25)	57.08 (3.20)	60.42 (3.16)	0.46
Race: White (%)	56.67 (2.26)	60.00 (3.17)	53.33 (3.23)	0.14
Race: Black (%)	31.46 (2.12)	29.58 (2.95)	33.33 (3.05)	0.38
Race: Other (%)	11.88 (1.48)	10.42 (1.98)	13.33 (2.20)	0.32
Hispanic (%)	9.79 (1.36)	8.33 (1.79)	11.25 (2.04)	0.28
College (%)	53.33 (2.28)	55.83 (3.21)	50.83 (3.23)	0.27
Married (%)	43.96 (2.27)	45.00 (3.22)	42.92 (3.20)	0.65
Employed (%)	69.17 (2.11)	70.83 (2.94)	67.50 (3.03)	0.43
Income	87727.27 (3483.28)	92060.09 (5396.21)	83318.78 (4379.17)	0.21
Homeowner (%)	49.38 (2.28)	48.75 (3.23)	50.00 (3.23)	0.78
Democrat (%)	52.92 (2.28)	54.58 (3.22)	51.25 (3.23)	0.47
MPL Starting Value was 21 (&)	33.12 (2.15)	34.17 (3.07)	32.08 (3.02)	0.63
MPL Starting Value was 46 (&)	34.17 (2.17)	32.50 (3.03)	35.83 (3.10)	0.44
MPL Starting Value was 81 (&)	32.71 (2.14)	33.33 (3.05)	32.08 (3.02)	0.77
Observations	480	240	240	

Notes: This table reports reduced-form results using the contingent valuation survey. Column (1) presents the overall mean. Column (2) presents the mean for the control group in the list experiment. Columns (3) presents the mean for the treatment group in the list experiment. Column (6) is a p-value from an F-test of joint equality across treatment conditions.

Table A.3: Balance Tests

	(1) HEART	(2) Post X Qualify
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 $Qualify$. 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 refer to tests of 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.4: Tests of Monotonicity Assumption

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Main Sample	Subsample Sh. Asian Above p50	Subsample Sh. Black Above p50	Subsample Sh. White Above p50	Subsample Sh. Hispanic Above p50	Subsample Unemp. Rate Above p50	Subsample Unemp. Rate Below p50
Post X Qualify	0.147^{***}	0.137^{***}	0.139^{***}	0.147^{***}	0.142^{***}	0.157^{***}	0.137^{***}
	(0.00340)	(0.00392)	(0.00357)	(0.00568)	(0.00431)	(0.00511)	(0.00413)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Call Type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	161054	82419	78482	79624	80528	78807	82247

Notes: This table shows the results from first-stage regressions estimating the probability of a HEART response using the implementation of HEART as an instrument for different subsamples. The interaction between the time since introduction and an indicator for calls qualifying for a HEART response serves as the instrument, with unqualified calls as the control group. The regression controls for the covariates in Table A.3 and includes fixed effects for day, hour, priority, month–year, and beat. Standard errors are clustered at the address level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.5: Main Specification vs. TSLS Results

	(1) Mean Pre-HEART	(2) Main Analysis Estimate	(3) Main Analysis SE	(4) TSLS Estimate	(5) TSLS SE
A) Crime					
Crime Report	0.161	-0.097	0.004	-0.079	0.024
Violent Crime	0.078	-0.031	0.002	-0.086	0.014
Property Crime	0.013	-0.004	0.001	-0.032	0.008
Other Crime	0.093	-0.065	0.003	-0.021	0.020
B) Arrest					
Arrest Report	0.048	-0.024	0.002	-0.057	0.013
Arrest for Violent Crime	0.028	-0.009	0.001	-0.035	0.008
Arrest for Property Crime	0.006	-0.002	0.000	-0.020	0.004
Arrest for Other Crime	0.029	-0.018	0.002	-0.032	0.011
C) Other Outcomes					
Need Back-Up	0.025	0.043	0.002	0.022	0.010
Total Time (min.)	50.017	-2.592	0.901	-21.556	5.117
Dispatch Time (min.)	4.941	-2.313	0.212	-1.105	1.079
Clearing Time (min.)	45.076	-0.279	0.794	-20.450	4.974
Use of Force (X 100)	0.130	-0.010	0.033	-0.266	0.209
Any New Call	0.960	-0.010	0.004	0.182	0.029
New Call with Violence	0.034	-0.002	0.001	-0.029	0.012

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, and odd-numbered columns display the corresponding standard errors. The table includes results from the main DiD specification and TSLS. All regressions account for day-of-week, hour, priority, month–year, and beat fixed effects. Standard errors are clustered at the address level.

Table A.6: First-Stage Estimates for HEART Response Types

	(1) Civilian Phone	(2) Civilian In-Person	(3) Co-Response
Instr. Civ. Phone	0.172***	0.0513***	0.0169***
	(0.0215)	(0.00793)	(0.00332)
Instr. Civ. In-Person	-0.0426***	0.137***	0.00776***
	(0.00455)	(0.00319)	(0.00134)
Instr. Co-Response	-0.0183***	-0.0999***	0.0252***
	(0.00141)	(0.00312)	(0.00126)
Mean Dep. Post-HEART	0.19	0.14	0.04
Time FE	Yes	Yes	Yes
Responder-Type Qual. FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
N	161054	161054	161054

Notes: This table reports estimates from first-stage regressions predicting the probability of receiving each HEART response type— civilian phone, civilian in-person, and co-response —using the implementation of HEART as an instrument. The instruments are defined as the interaction between an indicator for post-introduction periods and response-type-specific eligibility, with unqualified calls serving as the control group. Regressions control for the covariates listed in Table A.3 and include fixed effects for day-of-week, hour-of-day, call priority, calendar month–year, and geographic beat. Standard errors are clustered at the address level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.7: TSLS Estimates: Impact of HEART Response Types on Call Outcomes

	(1) Crime Report	(2) Arrest	(3) Total Time	(4) Need Backup
Civilian Phone	-0.185***	-0.00259	-58.95***	0.0622
	(0.0457)	(0.0178)	(9.130)	(0.0408)
Civilian In-Person	-0.0672**	0.0122	9.590*	0.0585***
	(0.0262)	(0.0169)	(5.325)	(0.0119)
Co-Response	0.173	-0.0990	-126.2***	0.0290
	(0.132)	(0.0659)	(27.73)	(0.0781)
F-Stat	143.8	143.8	143.8	143.8
Mean Dep. Civ. Phone	0.0595	0.00104	55.39	0.0449
Mean Dep. Civ. In-Person	0.120	0.0285	41.82	0.0255
Mean Dep. Co-Resp	0.194	0.0534	53.53	0.0272
Time FE	Yes	Yes	Yes	Yes
Responder-Type Qual. FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N	161054	161054	161054	161054

Notes: This table presents TSLS estimates of the impact of each HEART response type— civilian phone, civilian in-person, and co-response—on call-level outcomes, using response-type-specific eligibility interacted with post-period indicators as instruments. Each column reports results from a separate regression where the dependent variable is indicated at the top. All models control for the covariates listed in Table A.3 and include fixed effects for day-of-week, hour-of-day, call priority, calendar month–year, and geographic beat. Standard errors are clustered at the address level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.8: 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
Assault	$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
Driving 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
Weapons 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 Nontraffic 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 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.